PUBLISHED FOR THE BROOKINGS INSTITUTION BY

JOHNS HOPKINS UNIVERSITY PRESS

Brookings Papers ON ECONOMIC ACTIVITY

BPEA
FALL
2022

BROOKINGS

Janice Eberly and James H. Stock, Editors

Brookings Papers
ON ECONOMIC ACTIVITY

FALL 2022

Brookings Papers
ON ECONOMIC ACTIVITY

FALL 2022

JANICE EBERLY
JAMES H. STOCK
Editors

PUBLISHED FOR THE BROOKINGS INSTITUTION BY
JOHNS HOPKINS UNIVERSITY PRESS

THE BROOKINGS INSTITUTION
1775 Massachusetts Avenue, N.W., Washington, DC 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

FALL 2022

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. 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.

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. The general discussions are based on the working drafts presented at the conference, not the final versions published in the journal. Working drafts and video recordings of the Fall 2022 *BPEA* conference can be accessed at https://www.brookings.edu/events/bpea-fall-2022-conference/.

ACKNOWLEDGMENT We would like to thank the supporters of the *BPEA* conference and journal, including the Alfred P. Sloan Foundation; Brevan Howard Research Services; General Motors Company; the National Science Foundation, under grant no. 2048708; and State Farm Mutual Automobile Insurance Company. We gratefully acknowledge Dina Axelrad Perry for establishing the George L. Perry and William C. Brainard *BPEA* Chair.

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

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

Sanjay Patnaik *Brookings Institution*
George Perry *Brookings Institution*
Jonathan Pingle *UBS*
Eswar Prasad *Cornell University & Brookings Institution*
Brian Prest *Resources for the Future*
Valerie Ramey *University of California, San Diego*
Alessandro Rebucci *Johns Hopkins University*
Tristan Reed *World Bank*
David Reifschneider *formerly of the Federal Reserve Board of Governors*
Ricardo Reis *London School of Economics*
Giovanni Ricco *University of Warwick*
John Roberts *formerly of the Federal Reserve Board of Governors*
Matthew Rognlie *Northwestern University*
David Romer *University of California, Berkeley*
John Sabelhaus *Brookings Institution*
Ayşegül Şahin *University of Texas at Austin*
Claudia Sahm *Sahm Consulting*
Jake Schild *Bureau of Labor Statistics*
Moritz Schularick *University of Bonn*
Sam Schulhofer-Wohl *Federal Reserve Bank of Dallas*
Amit Seru *Stanford University & Hoover Institution*
Louise Sheiner *Brookings Institution*
Daniel E. Sichel *Wellesley College*
Chris Sims *Princeton University*
Tara Sinclair *George Washington University*
Mark Steinmeyer *Smith Richardson Foundation*
Betsey Stevenson *University of Michigan*
James Stock *Harvard University*
Michael Strain *American Enterprise Institute*
Phillip Swagel *Congressional Budget Office*
Daniel Tarullo *Harvard Law School*
Stijn Van Nieuwerburgh *Columbia Business School*
Stan Veuger *American Enterprise Institute*
Annette Vissing-Jorgensen *Federal Reserve Board of Governors*
Alice Volz *Federal Reserve Board of Governors*
William Wascher *Federal Reserve Board of Governors*
Tara Watson *Brookings Institution*
David Wessel *Brookings Institution*
Jon Willis *Federal Reserve Bank of Atlanta*
Justin Wolfers *University of Michigan*
Susan Woodward *Sand Hill Econometrics*
Abigail Wozniak *Federal Reserve Bank of Minneapolis*
Jonathan H. Wright *Johns Hopkins University*
Mindy Z. Xiaolan *University of Texas at Austin McCombs School of Business*
Pablo Zarate *Princeton University*
Haonan Zhou *Princeton University*

Karina Aguilar *Brookings Institution*
Janina Broker *Brookings Institution*
Haowen Chen *Brookings Institution*
Beyond Deng *Brookings Institution*
Siobhan Drummond *Brookings Institution*
Julia Paris *Brookings Institution*
David Skidmore *Brookings Institution*
Lorae Stojanovic *Brookings Institution*
Rayan Sud *Brookings Institution*

LAURENCE BALL
Johns Hopkins University

DANIEL LEIGH
International Monetary Fund

PRACHI MISHRA
International Monetary Fund

Understanding US Inflation during the COVID-19 Era

ABSTRACT This paper analyzes the dramatic rise in US inflation since 2020, which we decompose into a rise in core inflation as measured by the weighted median inflation rate and deviations of headline inflation from core. We explain the rise in core inflation with two factors: the tightening of the labor market as captured by the ratio of job vacancies to unemployment, and the pass-through into core inflation from past shocks to headline inflation. The headline shocks themselves are explained largely by increases in energy prices and by supply chain problems as captured by backlogs of orders for goods and services. Looking forward, we simulate the future path of inflation for alternative paths of the unemployment rate, focusing on the projections of Federal Reserve policymakers in which unemployment rises only modestly to 4.4 percent. We find that this unemployment path returns inflation to near the Federal Reserve's target only under optimistic assumptions about both inflation expectations and the Beveridge curve relating the unemployment and vacancy rates. Under less benign assumptions about these factors, the inflation rate remains well above target unless unemployment rises by more than the Federal Reserve projects.

Conflict of Interest Disclosure: Daniel Leigh and Prachi Mishra are employees of the International Monetary Fund, which conducts a review of all externally published pieces. 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, Fall 2022: 1–54 © 2023 The Brookings Institution.

After four decades of low US inflation, high inflation has emerged as a central economic problem of the COVID-19 era. As of September 2022, the rate of Consumer Price Index (CPI) inflation over the previous twelve months was 8.2 percent.[1] This experience has produced an outpouring of analyses of why inflation has risen and where it might be heading in the future. This paper seeks to contribute to this debate.

A central feature of our analysis is that we decompose the headline inflation rate into two components that are determined by different factors: core inflation and deviations of headline from core. We seek to explain core inflation with long-term expected inflation and the level of slack or tightness in the labor market, and to explain the noncore component of headline inflation with large price changes in particular industries. We also study the pass-through over time from these industry price shocks to core inflation, which can occur through the effects of headline inflation on wages and other costs of production.

Section I of this paper describes how we measure core inflation. Our primary measure is the weighted median inflation rate published by the Federal Reserve Bank of Cleveland, which strips out the effects of unusually large price changes in certain industries. This variable isolates the core component of inflation more effectively than the traditional core measure of inflation excluding food and energy prices, especially during the COVID-19 era, when much volatility in headline inflation has come from price changes in industries other than food and energy. In September 2022, weighted median inflation accounted for 7 percentage points of the 8.2 percent headline inflation rate.[2]

Section II studies the behavior of core inflation. A key feature of the analysis is that, following recent studies such as Furman and Powell (2021) and Barnichon, Oliveira, and Shapiro (2021), we measure the tightness of the labor market with the ratio of job vacancies (V) to unemployment (U). We find that the very high levels of V/U over 2021–2022 can explain much of the rise in monthly core inflation, especially during 2022. The rest of the rise is explained by a substantial pass-through of headline inflation shocks into core inflation.

These results help us understand why persistently high inflation has been a surprise to many economists—including us (Spilimbergo and

1. US Bureau of Labor Statistics, "Consumer Price Index News Release," https://www.bls.gov/news.release/archives/cpi_10132022.htm.

2. Federal Reserve Bank of Cleveland, "Median CPI," table "% Change Past 12 Months," https://www.clevelandfed.org/indicators-and-data/median-cpi.

others 2021)—who dismissed the run-up in inflation in mid-2021 as transitory. These economists typically measured labor market tightness with the unemployment rate, which has only fallen to but not below pre-pandemic levels, and they ignored the pass-through effect that can propagate the effects of headline inflation shocks.

Section III studies the pandemic-era shocks to headline inflation—the deviations of headline from core—that have contributed to inflation both directly and through the pass-through to core. We find that three factors have been most important in explaining this component of inflation: changes in energy prices; a measure of backlogs of goods and services orders from the information services firm IHS Markit Economics, which we believe captures the widely reported problems with supply chains; and changes in prices in auto-related industries.

Section III also performs a decomposition of the 6.9 percentage point rise in headline inflation between the end of 2020 and September 2022 (from 1.3 percent to 8.2 percent). It concludes that the combination of direct and pass-through effects from headline inflation shocks accounts for about 4.6 percentage points of the rise in twelve-month inflation. A rise in expected inflation accounts for 0.5 percentage points, and the rise in labor market tightness (measured by the ratio of vacancies to unemployment) accounts for 2 percentage points.

After analyzing the inflation experience to date, we turn to what might happen in the future. We focus on the question of what costs must be incurred for the Federal Reserve to meet its goal of reining in inflation. Federal Reserve officials have predicted a soft landing in which inflation returns to their target with only a modest increase in unemployment, while pessimists such as Lawrence Summers believe that disinflation will require a painful recession with high unemployment (Mellor 2022). Which outcome is more likely?

In our view, the answer depends largely on two factors, which we discuss in section IV. One is the relationship between unemployment and vacancies—the Beveridge curve. This relationship has shifted unfavorably during the pandemic: a given level of vacancies implies a higher level of unemployment. The unemployment costs of reducing inflation will be substantial if this relationship now remains unchanged, but the costs will be lower if a normalization of the labor market moves the Beveridge curve back toward its pre-pandemic position.

The second factor concerns long-term inflation expectations. By various measures, these expectations have been well-anchored through most of the pandemic period, but they have shown hints of increasing during 2022. The costs of containing inflation will be greater if these hints turn into a significant upward trend in expected inflation. It is difficult to predict how expectations will evolve, but we try to shed light on the possibilities by estimating the response of survey measures of expectations to movements in actual inflation.

Section V presents simulations of future inflation under alternative assumptions about these issues and about the path that the unemployment rate will follow. One unemployment path that we consider is the one forecast by Federal Reserve policymakers in their September 2022 *Summary of Economic Projections* (*SEP*), which peaks at 4.4 percent in 2023 and 2024 (FOMC 2022, table 1). In this case, if we make quite optimistic assumptions about both the Beveridge curve and inflation expectations, the inflation rate falls to a level near the Federal Reserve's target by the end of 2024. For a range of other assumptions, however, inflation stays well above the target. All in all, it seems likely that policymakers will need to push unemployment higher than these *SEP* projections if they are determined to meet their inflation goal.

Research over the last two years has yielded many insights into the factors behind inflation, and we borrow a number of these ideas, as we discuss throughout the paper. We seek to synthesize much of the recent thinking about inflation in a way that allows a transparent analysis of the data, a quantification of the impact of different factors, and an informed analysis of where inflation may head in the future.

I. Headline and Core Inflation

Our framework for studying inflation is based on a common decomposition:

$$\text{headline inflation} = \text{core inflation} + \text{headline shocks}. \tag{1}$$

Core inflation is also known as underlying inflation. We interpret this variable as a relatively slow-moving component of inflation that depends on inflation expectations and slack in the aggregate labor market, as in the textbook Phillips curve. Headline shocks—the deviations from core—are high-frequency movements arising from large price changes in particular sectors of the economy. Fluctuations in energy prices are a perennial source of headline shocks. During the pandemic, large price changes have also occurred in industries affected by shutdowns and supply disruptions, such as travel-related industries and used cars.

Here we describe how we measure core inflation and then examine the paths of headline and core inflation since 2020.

I.A. Measuring Core Inflation

The traditional measure of core inflation, the one that the Federal Reserve focuses on, is the inflation rate excluding food and energy prices (XFE inflation). This measure is so common that some economists use the term "core inflation" as a synonym for XFE inflation. However, a growing body of research argues that XFE inflation is a flawed measure of the economic concept of core inflation. The XFE measure was developed in the 1970s, when changes in food and energy prices caused large fluctuations in headline inflation (Gordon 1975). Since that time, volatility in headline inflation has also arisen from large price swings in industries besides food and energy, which are not filtered out of XFE inflation, and this phenomenon has been especially pronounced during the pandemic (Dolmas 2005; Ball and others 2021).

The shortcomings of the XFE core measure have led researchers to develop a class of alternatives: outlier exclusion measures that systematically filter out large price changes in *any* industry. These measures are weighted medians or trimmed means of the distribution of industry price changes. A number of studies find that these core measures are less volatile and more closely related to economic slack than XFE inflation (Dolmas and Koenig 2019; Verbrugge 2021; Ball and others 2021).[3]

This paper focuses on one specific outlier exclusion measure of core inflation, the weighted median CPI inflation rate published by the Federal Reserve Bank of Cleveland. It is the oldest such measure, published since the 1990s, and arguably the simplest. The online appendix considers other outlier exclusion core measures, such as the trimmed mean personal consumption expenditures (PCE) deflator inflation rate published by the Federal Reserve Bank of Dallas, and the weighted median PCE deflator inflation rate published by the Federal Reserve Bank of Cleveland.

With core inflation measured by weighted median inflation, we define headline inflation shocks as deviations of headline from median. By construction, our measures of core inflation and headline shocks sum to headline inflation.

Bryan and Cecchetti (1994) discuss the rationale for outlier exclusion measures of core inflation. In their framework, a large change in a sector's relative price affects the aggregate price level because, with costs of nominal price adjustment, large shocks to optimal prices have disproportionately

3. Similar evidence led the Bank of Canada to adopt a weighted median and trimmed mean as official measures of core inflation in 2016, replacing its CPIX measure, which is similar to XFE.

large effects on actual price changes. Removing outliers from the price distribution filters out the effects of relative price changes, thereby isolating the part of inflation determined by macroeconomic forces.[4]

The theory of core inflation has not been perfected, and more research is warranted. That said, in judging core inflation measures for present purposes, we believe that the proof of the pudding is in the eating. Throughout this paper, we find that our decomposition of headline inflation into median and deviations from median is a fruitful framework for understanding COVID-19-era inflation. We also show that much of our analysis would be infeasible if we measured core with XFE inflation.[5]

I.B. Headline and Core Inflation since 2020

We focus here on inflation in the CPI; the online appendix considers the PCE deflator. Figure 1 shows the paths of headline and median CPI inflation from January 2020 through September 2022 (the latest data available as this paper is written). Panel A shows monthly inflation at seasonally adjusted annualized rates, and panel B shows inflation over the past twelve months, a statistic that is widely reported in the media.

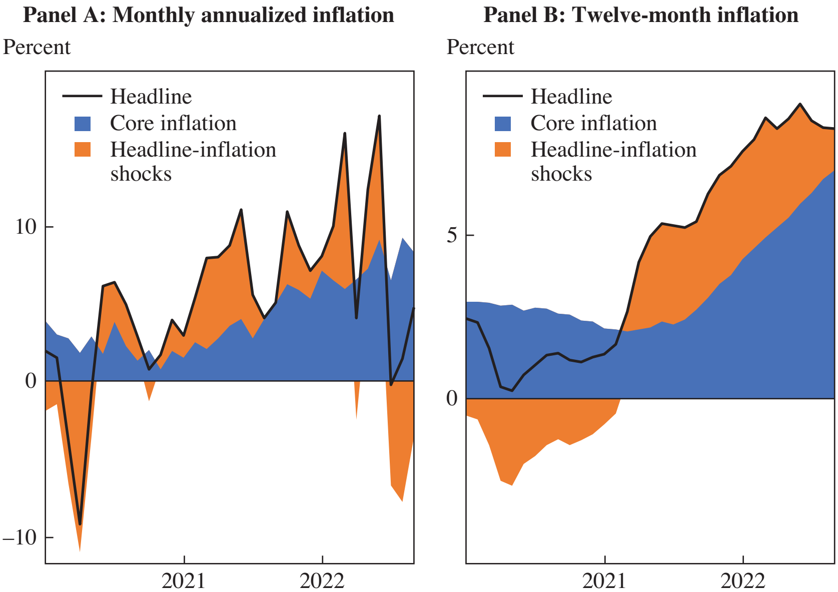
We can see from figure 1 that monthly headline inflation has been highly volatile, plunging close to –10 percent in April 2020, fluctuating up and down for the rest of that year, and coming in at 10 percent or higher at a number of points in 2021 and 2022. Monthly headline inflation soared to 17.1 percent in June 2022 and then fell to –0.2 percent in July, and it was 4.7 percent in September. The preponderance of high monthly readings in 2021 and the first half of 2022 pushed twelve-month headline inflation up to a peak (so far) of 9.1 percent in June 2022, and it was 8.2 percent in September.

Median inflation has been much less volatile, with the monthly series never changing by more than 3 percentage points from one month to the next. Median inflation drifted down in the first part of the pandemic, and as late as September 2021 the twelve-month median was still below its level in January 2020. This experience, and the common view that noncore inflation movements are transitory, helps explain the insouciance about inflation among many economists when a handful, such as Blanchard (2021) and

4. See Ball and Mazumder (2011) for more on these ideas.

5. Some economists criticize the weighted median on the grounds that the median industry is often related to housing, either rents or one of the four regional price indexes for owner-equivalent rent. However, it is not clear why it should matter which industry is the median or how much that varies.

Figure 1. CPI Inflation: Headline, Core, and Headline Inflation Shocks, 2020–2022



Sources: Federal Reserve Bank of Cleveland; authors' calculations.
Note: Core inflation is the weighted median CPI inflation rate from the Federal Reserve Bank of Cleveland.

Summers (2021), were first sounding an alarm. Since the middle of 2021, however, high monthly rates have led the twelve-month median to follow headline inflation upward, and it reached 7 percent in September 2022.

The following sections of the paper seek to explain this experience.

II. Explaining Core Inflation

Our basic framework explains core inflation with three variables: expected inflation; the tightness of the labor market; and past headline inflation shocks. The first two are the variables in the textbook Phillips curve, and the third captures the pass-through of headline inflation into core inflation that may occur through wages or other costs of producing output, channels emphasized by economists such as Blanchard (2022) and di Giovanni and others (2022). In our primary specification, expected inflation is measured by ten-year forecasts from the Survey of Professional Forecasters (SPF), labor-market tightness by the average ratio of job vacancies to unemployment (*V/U*) over the current and previous eleven months, and past headline shocks by the average deviation of headline from median inflation over the current and previous eleven months.

Our findings include the following:

- A core inflation equation estimated with pre-pandemic data provides a good fit to the path of core inflation during the pandemic. The increase in core inflation during 2021 and 2022 is explained by a combination of a rise in the *V/U* ratio to unprecedented levels and pass-through from adverse headline shocks, with the role of *V/U* increasing over the last year.
- There is some evidence of nonlinearity in the effect of *V/U* on core inflation, with a large positive marginal effect when *V/U* is either above or below its usual range. There is also strong evidence of asymmetry in the pass-through effects of headline shocks, which we find are negligible for shocks that reduce headline inflation but strong for shocks that increase headline inflation.
- We can estimate the contribution of the American Rescue Plan Act of 2021 (ARP) to core inflation using estimates from Barnichon, Oliveira, and Shapiro (2021) of the ARP's effects on *V/U*. For September 2022, we find a large effect on annualized monthly inflation of 4.2 percentage points. The effect on twelve-month inflation is 1.9 percentage points and rising.
- We find that both *V/U* and past headline shocks have strong effects on nominal wage growth. These results confirm the common view that labor market tightness and headline shocks transmit into core inflation through wage adjustment.

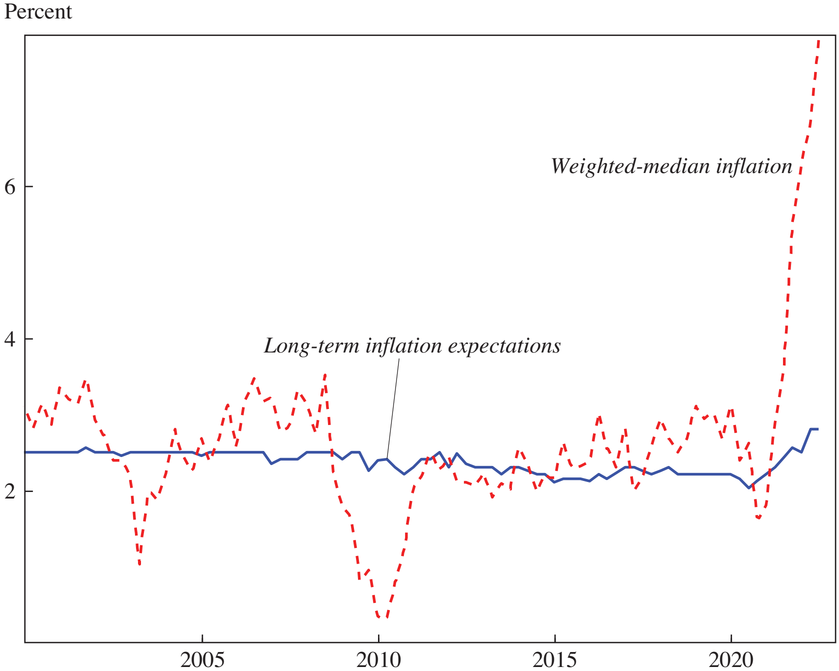
II.A. The Role of Expected Inflation

A central tenet of mainstream macroeconomics is that the inflation rate depends strongly on expected inflation. Following studies such as Hazell and others (2022) and our own past work, we measure expected inflation with the median ten-year-ahead CPI inflation forecast from the SPF. The online appendix considers another common measure, the five-year forecast from the University of Michigan Survey of Consumers.

For the period since 2000, figure 2 shows the path of the SPF expected inflation measure along with median inflation at the quarterly frequency.[6] We see that expected inflation has been stable. During 2000–2019 expected

6. Quarterly median inflation is constructed by aggregating monthly medians as described in Ball and Mazumder (2011).

Figure 2. Long-Term CPI Inflation Expectations and Median CPI Inflation, 2000–2022



Source: Survey of Professional Forecasters.
Note: Ten-year-ahead CPI inflation forecasts.

inflation averaged 2.36 percent, never deviating by more than 0.3 percentage points from this level. Economists have interpreted this level of CPI inflation as consistent with the Federal Reserve's 2 percent target for PCE deflator inflation, given the systematic tendency of CPI inflation to exceed PCE inflation by several tenths of a point (the average gap is 0.3 percentage points during 2009–2019 as reported on the Federal Reserve Bank of Atlanta's Underlying Inflation Dashboard).[7] These data support the common view that expected inflation has been well-anchored over the past two decades (Yellen 2019).

That said, there has been some increase in expected inflation during the pandemic, from 2.2 percent in 2019:Q4 to 2.8 percent in 2022:Q3. This rise presumably reflects the high realizations of actual inflation during this

7. Federal Reserve Bank of Atlanta, "Underlying Inflation Dashboard," https://www.atlantafed.org/research/inflationproject/underlying-inflation-dashboard.

period. A vital question, which we discuss in section IV, is whether expected inflation will de-anchor to a larger degree in the future.

In our econometric work, we assume that core inflation responds one-for-one to movements in long-run expected inflation.[8] The dependent variable in our equations is the difference between core inflation and expected inflation, which we call the "core inflation gap." We seek to explain this variable with labor market tightness and pass-through from headline inflation shocks.

II.B. The Effects of Labor Market Tightness, as Measured by V/U

Economists have long sought to explain short-run movements in the inflation rate with the level of tightness or slack in the labor market. Since Phillips (1958), the standard measure of labor market tightness has been the unemployment rate. To be sure, economists have developed more sophisticated measures that account for job vacancies and factors such as the search intensity of job seekers and firms (Abraham, Haltiwanger, and Rendell 2020) and hours of work of the employed (Faberman and others 2020). But up until the pandemic, the unemployment rate remained the most common measure of labor market tightness, in part because of its simplicity.

An important development in the last two years is that a number of inflation researchers, including Furman and Powell (2021), Barnichon and Shapiro (2022), and Domash and Summers (2022), have adopted the ratio of job vacancies to unemployment (V/U) rather than the unemployment rate as a simple measure of labor market tightness. This has been possible because of the data on vacancies collected since 2001 in the Bureau of Labor Statistics (BLS) Job Openings and Labor Turnover Survey (JOLTS), and Barnichon's (2010) extension of these data back to the 1950s using the help wanted index from the Conference Board. We follow this approach.[9]

The V/U ratio has strong theoretical appeal as a measure of wage pressures that feed into price inflation: V/U determines the threat points of the workers and firms that bargain over wages in search models (Mortensen

8. A Phillips curve specification where changes in long-run inflation expectations affect current inflation one-for-one is derived by Hazell and others (2022) in a New Keynesian framework under the assumption that shocks to the natural rate of unemployment and cost-push shocks are transitory. The authors show that under such conditions, long-run inflation expectations enter the Phillips curve with a coefficient of one.

9. Long ago, Medoff and Abraham (1982) argued that the job vacancy rate was a better measure of labor market tightness than the unemployment rate. But that paper did not have much impact on the Phillips curve literature, a likely reason being the poor quality of vacancy data before the JOLTS survey.

and Pissarides 1999). In addition, there is some evidence from the pre-pandemic era that *V/U* outperforms the unemployment rate in explaining both wage and price inflation, although the difference is not crystal clear because the two series are highly correlated (see the studies cited above and the online appendix).

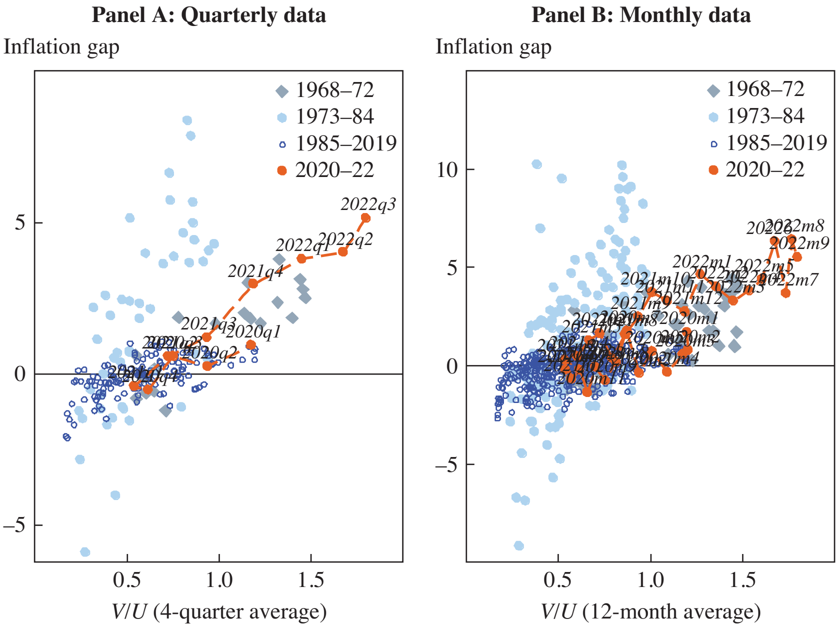
For the pandemic period, it is easier to distinguish the roles of unemployment and *V/U* because the two tightness measures have behaved differently. Over the first half of 2022, the unemployment rate averaged 3.7 percent, which is slightly above its January 2020 level (3.5 percent) and not far below the Congressional Budget Office's (CBO) estimate of the natural rate of unemployment (4.4 percent), so by that measure the labor market has not been especially tight.[10] In contrast, the average *V/U* ratio for the same period was 1.88, the highest it has been since 1951 when the Barnichon data begin. The recent levels of *V/U* imply a very tight labor market and potentially help explain the rise in inflation. (The divergence of the two tightness measures reflects a shift in the Beveridge curve relating unemployment and vacancies, which we analyze in section IV.)

COMPARING *V/U* AND THE INFLATION GAP We examine the relation between the inflation gap (median inflation minus expected inflation) and *V/U* in data back to 1968, when the Federal Reserve Bank of Cleveland's median series begins.[11] We examine both quarterly and monthly data and compare the current level of the inflation gap to an average of *V/U* over the current and previous three quarters or the current and previous eleven months. We use these averages as a parsimonious way of capturing the lags in the effects of labor market tightness that previous research typically finds. (As a robustness check, the online appendix considers the relation between the inflation gap and the current level of *V/U* alone.)

10. Bureau of Labor Statistics, "Labor Force Statistics from the Current Population Survey," LNS14000000, https://data.bls.gov/timeseries/LNS14000000; FRED Economic Data, Federal Reserve Bank of St. Louis, "Noncyclical Rate of Unemployment (NROU)," https://fred.stlouisfed.org/series/NROU.

11. Some data details: we splice the Federal Reserve Bank of Cleveland's old and new series for the median following Ball and Mazumder (2011). Data for long-term (ten-year-ahead) CPI inflation expectations come from the Federal Reserve Bank of Philadelphia website starting in 1979:Q4. These data come from the SPF starting in 1991:Q4 and from Blue Chip semiannual survey data from 1979:Q4 to 1991:Q1 (with interpolation in between surveys). For 1968:Q1 to 1979:Q3, we use forecasts from the data set on the Federal Reserve website, which are constructed from a mixture of surveys and econometric work. These forecasts are for PCE deflator inflation; we add 0.4 percentage points to obtain CPI inflation forecasts, following a rule of thumb that the Federal Reserve staff used in constructing the data set. In our monthly analysis, we use quarterly forecasts for the middle month of each quarter and interpolate between these months.

Figure 3. Inflation Gap versus Ratio of Vacancies to Unemployed, 1968–2022



Sources: Survey of Professional Forecasters; authors' calculations.
Note: "Inflation gap" is the difference between median and long-term expected inflation. Long-term expected inflation is the ten-year-ahead CPI inflation forecast. *V*/*U* denotes ratio of vacancies to unemployed (four-quarter or twelve-month average).

Figure 3 shows quarterly and monthly scatterplots of the inflation gap against the averages of *V/U*. We use different markers for the observations in four parts of the sample: 1968–1972; 1973–1984, the period of high inflation and then disinflation ushered in by the first oil shock; 1985–2019, a long period of low inflation that includes both the Great Moderation of 1985–2007 and the subsequent Great Recession and recovery; and the COVID-19 era of 2020–2022 (through 2022:Q3 or September).[12]

Notice first that the 1973–1984 period jumps out as one with unusually high inflation gaps and a steep relation between the gap and *V/U*. This anomalous behavior likely reflects the pre-Volcker monetary regime of large inflation shocks, accommodative policy, and unanchored expectations. The fluctuations in inflation are also magnified by the treatment

12. The observation for *V* in September 2022 is not available as this is written. We estimate it by simply assuming that *V* is the same in September as in August.

of housing in the CPI before 1982, which distorts inflation measurement relative to current practice when interest rates are volatile (Bolhuis, Cramer, and Summers 2022). In any case, we do not analyze this period further in this paper.

Starting in 1985, the data appear consistent with an upward-sloping relation between the inflation gap and *V/U* that is fairly stable. The observations for late 2021 and 2022 appear in the upper right of the graphs, with significantly higher gaps and a tighter labor market than at any previous time since 1985. The recent levels of the inflation gap appear roughly consistent with the unusually high levels of *V/U* and the pre-pandemic relation between the two variables.

The recent observations are also fairly consistent with those from the late 1960s, a period of overheating represented by the diamond shapes on the right sides of the graphs. That was the last period with levels of *V/U* comparable to 2021–2022, and the inflation gap reached similar levels. We believe this fact is noteworthy, although the econometric work in this paper will use only data starting in 1985 to address concerns that the structure of the economy was very different in the 1960s.

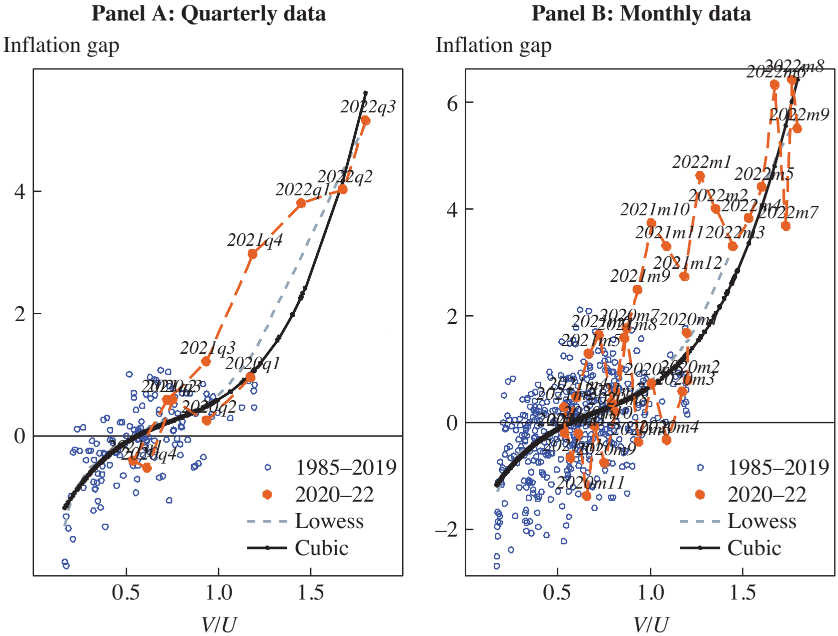
To aid in interpreting the scatterplots, figure 4 shows the results of fitting flexible curves to the data for 1985–2022. We consider a cubic function of *V/U* and a lowess estimator with a bandwidth of 0.8, which produce similar results. The data suggest a fairly flat relation for midrange levels of *V/U* and a steeper relation on either side: *V/U* has a larger marginal effect on core inflation when its level is unusually high or unusually low. The levels of the inflation gap are somewhat above the fitted curves in late 2021, but the most recent observations are close to the curves.

THE IMPORTANCE OF CORE MEASUREMENT One thing that distinguishes this paper from most inflation research is that we measure core inflation with the weighted median inflation rate. We can now see some evidence that this choice is important. Figure 5 repeats figure 3, the scatterplots of the inflation gap against *V/U*, but with core inflation measured in the traditional way with inflation excluding food and energy prices (XFE inflation). We see that the relation becomes noisier before the pandemic, and that during the pandemic XFE inflation fluctuates erratically with no clear relation to movements in *V/U*. These patterns reflect the noise in XFE inflation arising from large price changes in industries other than food and energy.

II.C. Pass-Through from Headline Inflation Shocks

Many studies of core inflation, whether measured by the weighted median or by XFE inflation, seek to explain its behavior with expected inflation and

Figure 4. Fitted Relationship: Inflation Gap versus Ratio of Vacancies to Unemployed, 1985–2022



Sources: Survey of Professional Forecasters; authors' calculations.

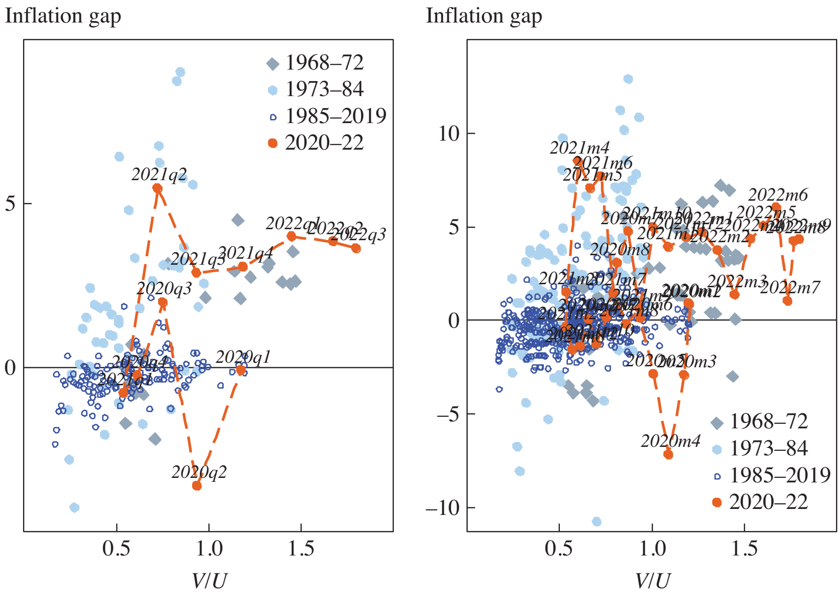
Note: "Lowess" denotes locally weighted scatterplot smoothing strategy for fitting a smooth curve to data points. "Inflation gap" is the difference between median and long-term expected inflation. Long-term expected inflation is the ten-year-ahead CPI inflation forecast. *V*/*U* denotes ratio of vacancies to unemployed (four-quarter or twelve-month average).

slack and assume implicitly that the evolution of core inflation is unrelated to the deviations of headline from core (Spilimbergo and others 2021). Policymakers sometimes suggest that headline shocks can be ignored in analyzing and forecasting core inflation. However, some strands of the literature call this view into question, arguing that shocks to headline inflation can eventually be passed through into core inflation.

One possible mechanism, stressed by researchers such as Blanchard (2022), is wage adjustment: increases in the cost of living as measured by headline inflation influence wage demands throughout the economy and thereby contribute to core inflation. Blanchard suggests that this effect may be especially strong for large movements in inflation, which are salient to wage setters. Another pass-through channel arises because the goods and services whose price changes contribute to headline shocks are inputs into

Figure 5. CPI Inflation Excluding Food and Energy versus Ratio of Vacancies to Unemployed, 1968–2022

Panel A: Quarterly data **Panel B: Monthly data**



Sources: Survey of Professional Forecasters; Bureau of Labor Statistics; authors' calculations.

Note: This figure repeats figure 3, the scatterplots of the inflation gap against *V*/*U*, but with core inflation measured in the traditional way with inflation excluding food and energy prices (XFE). "Inflation gap" is the difference between XFE inflation and long-term expected inflation. Long-term expected inflation is the ten-year-ahead CPI inflation forecast. *V*/*U* denotes ratio of vacancies to unemployed (four-quarter or twelve-month average).

the production of other goods, so the price changes affect costs of production. Research at the European Central Bank (2014) stresses this effect in analyzing the transmission of oil price shocks into inflation.

We explore the effects of headline shocks as captured by the average of headline inflation minus core inflation over the same four-quarter or twelve-month period over which we measure *V*/*U* in the analysis above. This approach is consistent with European Central Bank (2014) work on oil shocks, which finds that they transmit into inflation slowly. In the online appendix we experiment with headline shocks averaged over shorter periods and find that they do not explain core inflation as well.

Pass-through effects are potentially important in the pandemic era because headline shocks have been large. The twelve-month average of these shocks has risen as high as 3.7 percentage points (in March 2022), far higher than at any point since the 1970s, although it had fallen to 1.4 percentage points as of September 2022.

II.D. An Equation for Core Inflation

For the rest of this paper, we seek to explain the core inflation gap (median minus expected inflation) with four-quarter or twelve-month averages of *V/U* and headline shocks. We denote the headline shock variable by *H*. There are reasons to think that the effects of *V/U* and *H* may be nonlinear. For example, Blanchard (2022) emphasizes the salience of large shocks; Ball and Mankiw (1994) theorize that shocks have asymmetric effects in the presence of menu costs and trend inflation; and a number of studies find asymmetric pass-through effects from crude oil to retail fuel prices ("rockets and feathers").[13] Therefore, we allow for nonlinearities in a flexible way, by including cubic functions of *V/U* and *H* in the core inflation equation. Despite our nontraditional measure of labor market tightness, we call this relation the Phillips curve.

ESTIMATES Table 1 presents estimates of our Phillips curve. We report results for both quarterly and monthly data, which are similar. The data start in 1985, which is approximately the beginning of the Great Moderation period of low macroeconomic volatility (Bernanke 2004). We present estimates for the pre-pandemic period of 1985–2019 and also for that period extended to the present (2022:Q3 or September). We do not present results for the pandemic period alone, which would mean estimating seven parameters with eleven quarters of data.

For both samples, the squared and cubic terms are statistically significant for both *V/U* and *H*: the data indicate nonlinearity in the effects of these variables.[14] To aid in interpreting the results, figure 6 shows the shapes of the estimated cubic functions for monthly data from 1985 to the present, with 95 percent confidence intervals. We show the functions over the ranges of *V/U* and *H* in the data. Panel A shows the fitted values of the inflation gap as a function of *V/U* with the headline shock variable set to zero, which reveals a shape similar to that of the bivariate relation between the inflation gap and *V/U* in figure 4. Panel B shows the effect of *H* for a given *V/U*, which proves to be strikingly asymmetric: negative values of *H*, that is, headline inflation rates below median inflation, have negligible effects on

13. See, for example, Borenstein, Cameron, and Gilbert (1997) and Owyang and Vermann (2014).

14. In one case, monthly data for 1985–2019, the joint significance of the $(V/U)^2$ and $(V/U)^3$ terms is borderline ($p = 0.053$). These terms are strongly significant in quarterly data for the same period ($p = 0.012$) and in both quarterly and monthly data through 2022 ($p < 0.01$).

Table 1. Phillips Curve Estimates: Median CPI Inflation

	(1) Quarterly 1985–2019	*(2) Quarterly 1985–2022*	*(3) Monthly 1985–2019*	*(4) Monthly 1985–2022*
V/U	11.039***	9.024***	9.553**	9.140***
	(3.645)	(2.120)	(4.297)	(2.234)
V/U^2	−13.261**	−10.083***	−10.879*	−10.328***
	(5.485)	(2.383)	(6.435)	(2.545)
V/U^3	5.541**	4.032***	4.439	4.241***
	(2.530)	(0.789)	(2.958)	(0.863)
H	0.021	0.031	0.010	0.058
	(0.068)	(0.074)	(0.073)	(0.075)
H^2	0.155***	0.081***	0.128***	0.089***
	(0.041)	(0.016)	(0.035)	(0.019)
H^3	0.054***	0.026**	0.053***	0.031**
	(0.019)	(0.010)	(0.017)	(0.012)
Constant	−3.026***	−2.616***	−2.759***	−2.654***
	(0.747)	(0.557)	(0.879)	(0.586)
Observations	140	151	420	453
R^2	0.512	0.761	0.284	0.575
Adjusted R^2	0.490	0.751	0.274	0.569

Source: Authors' calculations.

Note: V/U denotes ratio of vacancies to unemployed (four-quarter or twelve-month average). H denotes headline inflation shock (four-quarter or twelve-month average). Newey-West standard errors with four lags (quarterly data) and twelve lags (monthly data) in parentheses.

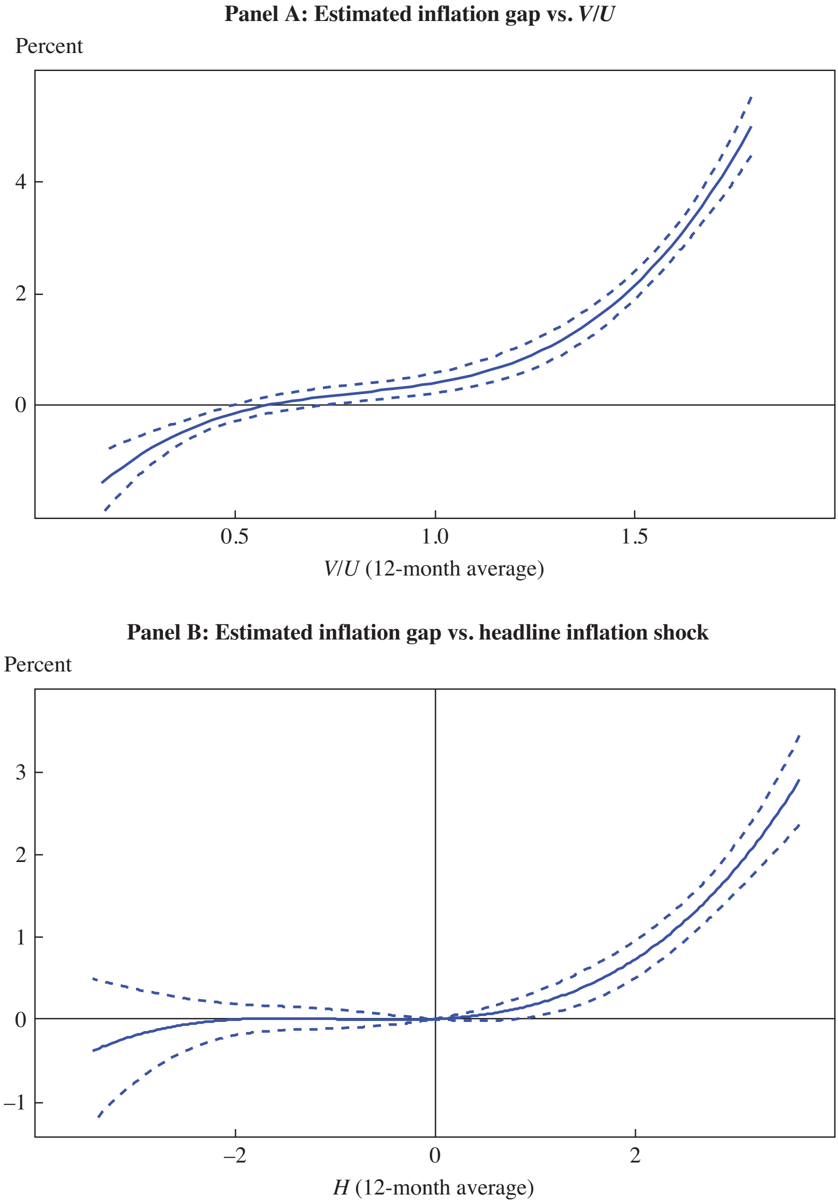
***$p < .01$, **$p < .05$, and *$p < .10$

core inflation, but positive values of H raise core inflation. Future research should explore the sources of this asymmetry.

EXPLAINING CORE INFLATION DURING THE PANDEMIC Do the variables in our inflation equation explain core inflation during the pandemic? To address this question, we compare actual and fitted values of the monthly core inflation gap from 2020 to the present in figure 7. Panel A presents results based on the full sample from 1985 to the present and panel B based on the pre-pandemic period from 1985 through 2019. In both cases, the fitted and actual values are close to each other. Note that panel B is an out-of-sample forecast; the good fit in this case means that we can explain the pandemic experience based on the paths of V/U and H and the estimated effects of these variables in the pre-pandemic period.

Figure 7 also shows the fitted values for the core inflation gap with the actual path of V/U but with the headline shock variable H set to zero. We interpret these paths as showing the contribution of labor market tightness to the rise in the inflation gap during the pandemic; the pass-through from headline shocks is the difference between these fitted values and those with

Figure 6. Estimated Inflation Gap as a Function of Slack and Headline Inflation Shocks, 1985–2022
(Percentage points; monthly data)

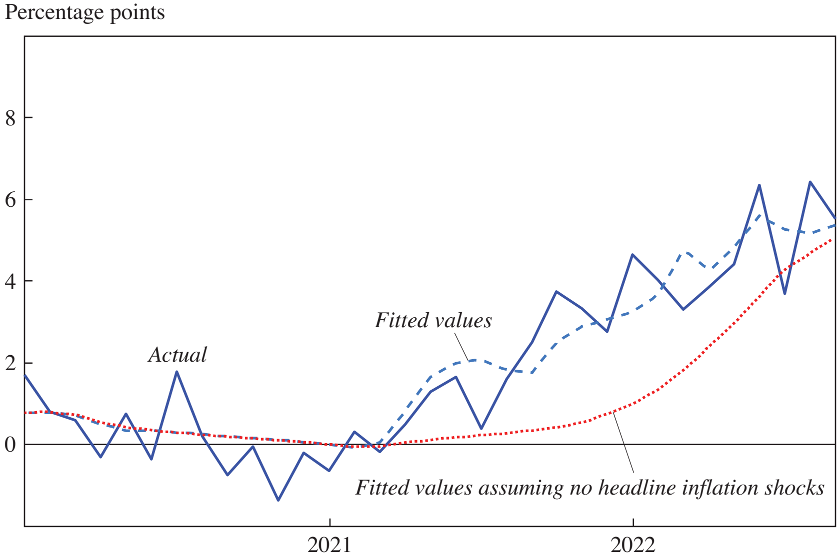


Sources: Survey of Professional Forecasters; authors' calculations.

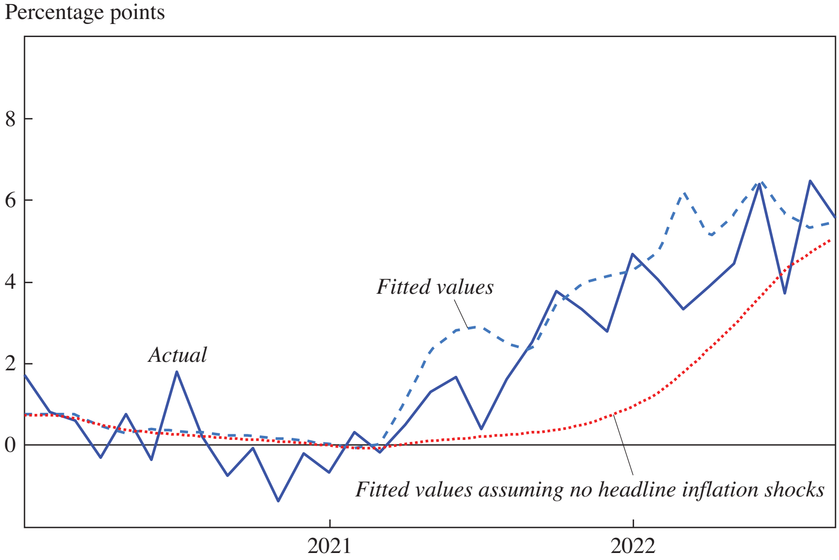
Note: Panel A reports fitted values for constant term and *V*/*U* terms from equation estimates reported in table 1 (column 4). Panel B reports fitted values for headline inflation shock (*H*) terms. Dotted lines indicate 95 percent confidence interval. Inflation gap denotes monthly annualized median CPI inflation minus long-term inflation expectations.

Figure 7. Predictions for Median Inflation Gap, 2020–2022

Panel A: Sample, 1985–2022



Panel B: Sample, 1985–2019



Sources: Survey of Professional Forecasters; authors' calculations.

Note: Figure reports fitted values from Phillips curve model estimated for the full sample (table 1, column 4) and for the pre-pandemic sample (table 1, column 3). Inflation gap denotes monthly annualized median CPI inflation minus long-term inflation expectations.

the actual path of H. We can see that the causes of rising core inflation have changed over time. Through most of 2021, there was little contribution from labor market tightness, but a strong pass-through effect pushed inflation up. In 2022, by contrast, the pass-through effect has diminished and the effect of labor market tightness has risen and become the main cause of high core inflation.

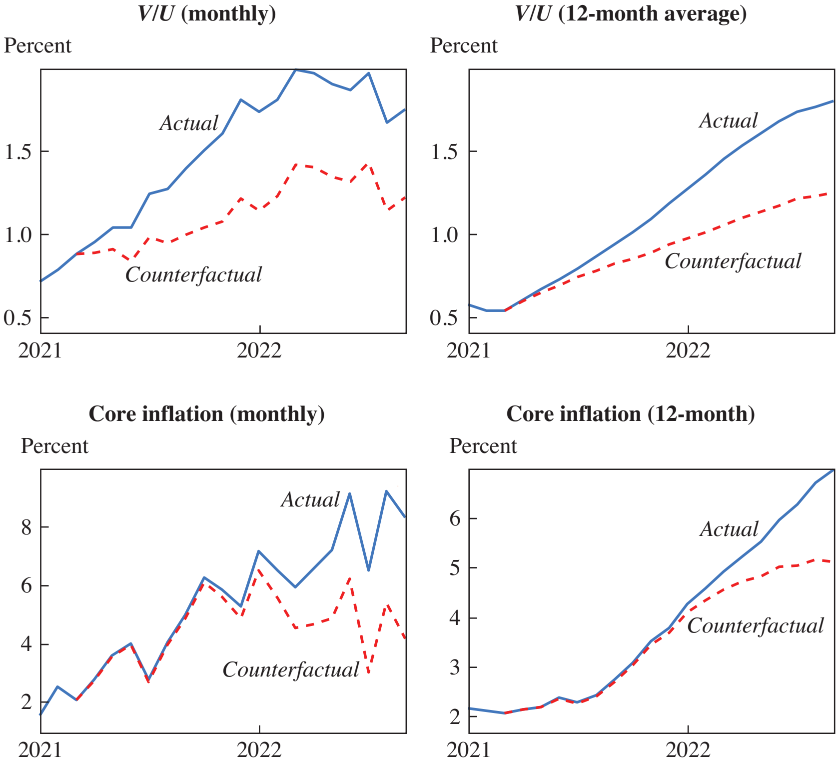
CORE INFLATION MEASUREMENT Once again, our choice of a core inflation measure is critical for our results. The online appendix reports a version of the regressions in table 1 with core inflation measured by XFE inflation. In this case, headline shocks are deviations of headline inflation from XFE inflation, which are determined by changes in the relative prices of food and energy. With these changes, we find almost no evidence of a pass-through from past headline shocks to core inflation. In addition, our core inflation equation estimated through 2019 fails to predict any rise in inflation during the pandemic era, in contrast to the equation's good performance when core is measured by weighted median inflation.

II.E. The Role of the American Rescue Plan

Many economists and politicians blame the American Rescue Plan Act (ARP) passed in March 2021—the $1.9 trillion Biden stimulus plan with enhanced unemployment benefits and stimulus checks—for the overheating of the economy and rise in inflation. Our framework suggests there was some such effect: to the extent the policy stimulated demand, it presumably reduced unemployment and increased vacancies, and the higher V/U ratio raised inflation. Here we seek to quantify this effect.

We do not estimate the effects of the ARP on the labor market; rather, we take estimates from a previous study by Barnichon, Oliveira, and Shapiro (2021) and then derive the implied effects on inflation. The Barnichon study is useful for our purpose because it directly estimates the effects of the ARP on the V/U ratio. It uses methodology from Ramey and Zubairy (2018) for estimating the effects of fiscal policy based on identifying changes in government spending related to wars or geopolitical events. A caveat is that the effects on V/U are uncertain because pandemic-era lockdowns could have reduced the response of consumption to changes in government spending (Seliski and others 2020). Barnichon, Oliveira, and Shapiro (2021) conclude that the ARP increased V/U by approximately 0.6 at the end of 2021 and 0.5 at the end of 2022. We obtain a monthly path for the effects by linearly interpolating between these values. Figure 8 shows the actual path of V/U over 2020–2022 and the path when we subtract the effects of the stimulus.

Figure 8. Counterfactual Scenario without the American Rescue Plan



Source: Authors' calculations.

Note: Data on the impact of the American Rescue Plan on *V/U* come from Barnichon, Oliveira, and Shapiro (2021). Core inflation denotes median CPI inflation. Monthly inflation is annualized. The impact on core inflation derived from the Phillips curve relation estimated for 1985–2022.

Using these results, we compare the actual path of core inflation to the path in the counterfactual without the ARP. The counterfactual path is computed by subtracting the effect of the *V/U* difference in the two cases, which we compute from the relation between *V/U* and the inflation gap shown in figure 6; we assume that expected inflation is unaffected so the effect on core inflation equals the effect on the gap. We find that the difference between the two inflation paths was small in 2021 but has risen greatly in 2022. In September 2022, monthly core (median CPI) inflation is 4.2 percentage points lower in the counterfactual (4.1 percent rather than 8.3 percent) and twelve-month core inflation is 1.9 percentage points lower (5.1 percent rather than 7.0 percent). This difference amounts to about

40 percent of the rise in twelve-month core inflation from the end of 2020 to September 2022 and about one-quarter of the rise in twelve-month headline inflation.

A caveat: we have assumed that labor market tightness is the only channel through which the ARP has affected inflation. Summers suggests that the overheating of the economy arising from the ARP has helped cause supply chain problems, which in our framework can contribute to the headline shock component of inflation (Summers and Zakaria 2022). To the extent that such effects are present, our estimate of the ARP's effects on inflation should be interpreted as a lower bound.

II.F. Wage Inflation

In arguing that labor market tightness and past headline shocks affect price inflation, many researchers suggest that the channels are through wages: wage inflation responds to V/U and H, and wage inflation increases firms' costs and therefore passes into price inflation. We examine these ideas with data on wage inflation as measured by the growth rate of the employment cost index, a quarterly measure commonly used in previous work.

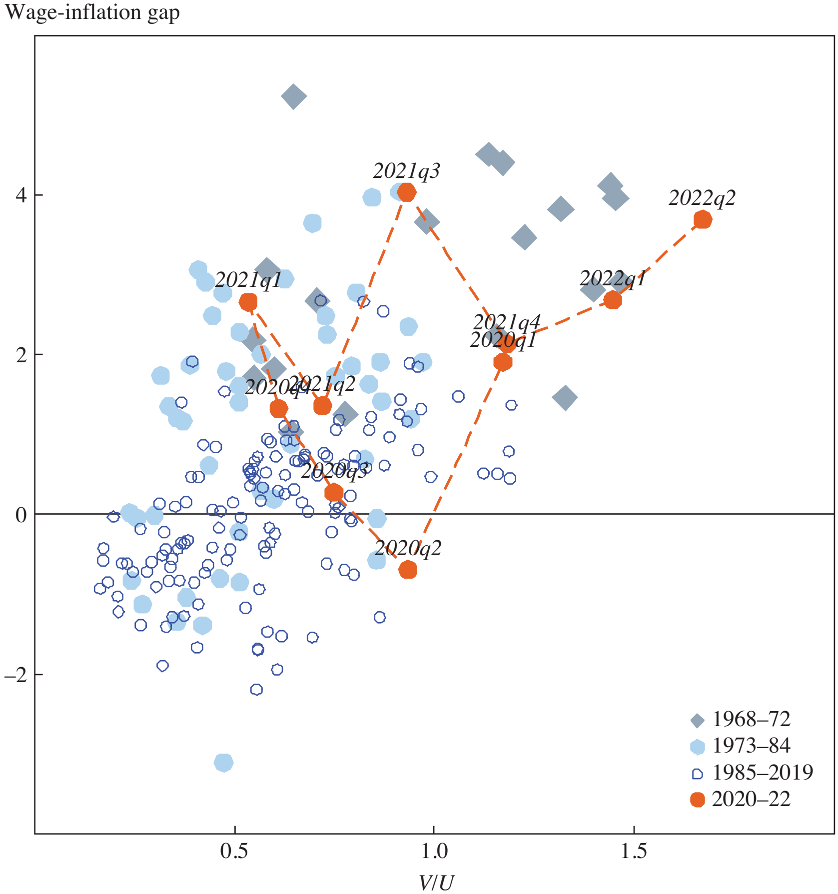
Figure 9 shows a scatterplot of the wage-inflation gap—wage inflation minus expected inflation—against the four-quarter average of V/U for the period 1968–2022:Q2. We see an upward-sloping relationship, albeit one that is somewhat noisy. The relationship appears consistent across time (here, the 1970s do not jump out as unusual).[15]

To examine wage behavior more carefully, we estimate versions of the Phillips curves in table 1 with the wage-inflation gap rather than median price inflation on the left side. We again include cubic functions of V/U and H, and following previous work on wage inflation we add a measure of trend productivity growth (output per hour in the nonfarm business sector smoothed with the Hodrick-Prescott filter with $\lambda = 16{,}000$). We present the estimated equations in the online appendix and focus here on the effects of V/U and H as captured in graphs.

For 1985–2022, figure 10 shows the wage-inflation gap as a function of V/U (with H set to zero and trend productivity set to its sample mean), and the effect of H, with 95 percent confidence intervals. For reference, we superimpose the relations between median price inflation and the two variables (estimated here with quarterly data). We find that the effects of

15. We leave out one big outlier: 1972:Q1, with an annualized wage increase of 13.2 percent. This increase may reflect the end of the Nixon administration's wage and price freeze.

Figure 9. Wage-Inflation Gap versus Ratio of Vacancies to Unemployed



Sources: Survey of Professional Forecasters; authors' calculations.

Note: "Wage-inflation gap" denotes the difference between quarterly wage inflation and long-term expected inflation. Long-term expected inflation is the ten-year-ahead CPI inflation forecast. *V/U* denotes the ratio of vacancies to unemployed (four-quarter average).

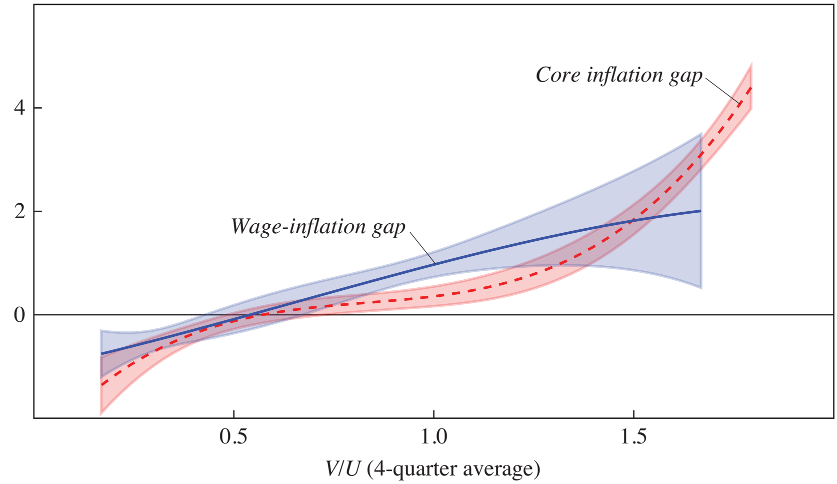
V/U and *H* on wage inflation are broadly similar to their effects on price inflation, consistent with the common view of transmission from wages to prices.

In contrast to our results for price inflation, the estimated effect of *V/U* on wage inflation is approximately linear. We are not sure whether this result reflects a meaningful difference between price and wage behavior, or simply the difficulty of detecting nonlinearities with noisy wage data.

Figure 10. Estimated Wage and Core Inflation Gaps as Functions of Slack and Headline Inflation Shocks, 1985–2022

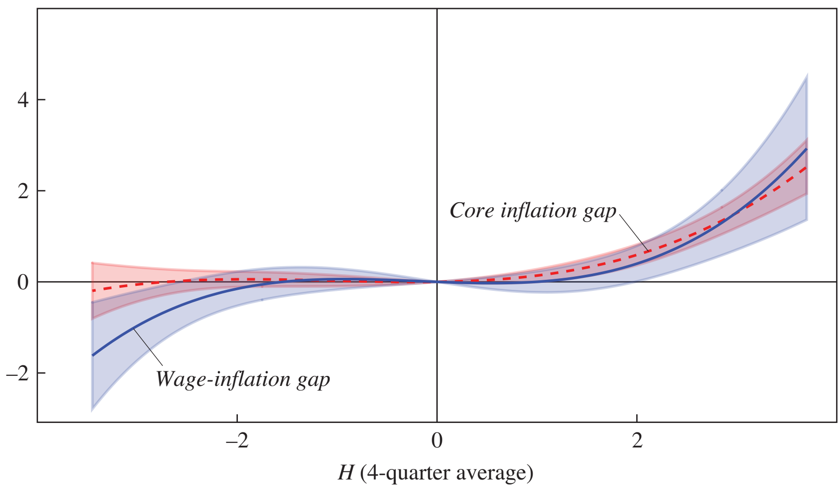
Panel A: Inflation gap vs. *V/U*

Percentage points, quarterly data



Panel B: Inflation gap vs. headline inflation shock

Percentage points, quarterly data



Sources: Survey of Professional Forecasters; authors' calculations.

Note: For price inflation, panel A reports fitted values for constant and ratio of vacancies to unemployed (*V/U*) terms based on specification reported in table 1 (column 2); panel B reports fitted values for headline inflation shock (*H*) terms. For wage inflation, fitted values for constant, *V/U*, and productivity growth terms are based on specifications reported in online appendix table 10 (column 2) with productivity growth set at its sample mean. Inflation gap denotes quarterly core (median) CPI inflation or wage inflation minus long-term inflation expectations. Bands (shaded areas) report 95 percent confidence interval.

III. Explaining Headline Inflation

We now examine the behavior of headline inflation, the variable that the public cares about. We first examine the causes of headline inflation shocks during the pandemic and find important roles for three variables: changes in energy prices, backlogs of orders for goods and services, and changes in auto-related prices. We then combine these results with those of the previous section to decompose the pandemic-era rise in inflation into the various factors that have influenced core inflation and headline shocks. Finally, we ask why many economists have been so surprised by the rise in inflation. Unanticipated shocks to the economy have played a role, but so have flaws in our pre-pandemic understanding of inflation drivers.

III.A. Explaining Headline Shocks

Here we seek to explain the monthly deviations of headline from core inflation, which affect inflation both directly and through their pass-through to core. These deviations arise from shocks that cause large price changes in certain sectors of the economy and thereby push the mean of the price change distribution (headline inflation) away from the median. These shocks can be shifts in either industry supply (such as disruptions in the supply of oil) or industry demand (such as the fall in demand for many services at the onset of the pandemic). Unlike many studies of inflation, we do not try to estimate the relative importance of supply and demand shocks.

Large shocks occur in different sectors of the economy at different times. (That is why our core inflation measure filters out all large price changes rather than excluding a fixed set of industries.) We seek to identify the sources of headline inflation shocks during the pandemic era—the light-shaded part of the inflation decomposition in figure 1.

We explore the possible roles of many variables that are cited in discussions of pandemic-era inflation. These variables include price changes in certain sectors of the economy, such as food and energy. They also include variables that have affected multiple sectors, such as measures of the severity of COVID-19 lockdowns and disruptions in production and distribution in the economy. Table 2 presents simple regressions of headline inflation shocks on each of these variables and multiple regressions on the variables that seem most important.

In the simple regressions, the variables with the most explanatory power are, in order of importance (with adjusted R^2 statistics in parentheses): energy price shocks, measured as energy price inflation minus median inflation (0.646); the IHS Markit Economics index of firms' backlogs of goods and

Table 2. Explaining Headline Inflation Shocks, 2020–2022
(Dependent variable: Headline–Median CPI monthly annualized inflation)

A. Bivariate regressions

	(1) Energy price inflation	(2) Food price inflation	(3) Harper Charter Rate	(4) Baltic Dry	(5) Supplier delivery times	(6) FRBNY Supply Chain Index	(7) Backlogs of work	(8) Goods share of real consumption	(9) Weighted average of car inflation rates	(10) COVID-19 intervention stringency
Coefficient	0.064***	−0.233	0.000	0.002***	−0.172***	0.887*	0.585***	2.115***	0.081***	0.033
	(0.013)	(0.274)	(0.001)	(0.001)	(0.056)	(0.459)	(0.114)	(0.640)	(0.021)	(0.042)
Observations	33	33	33	33	33	33	33	33	33	33
R^2	0.657	0.041	0.026	0.216	0.138	0.043	0.447	0.277	0.216	0.019
Adjusted R^2	0.646	0.0104	−0.00564	0.191	0.110	0.0125	0.429	0.253	0.191	−0.0127

B. Selected multivariate regressions

	(1)	(2)	(3)	(4)
Energy price inflation	0.051***	0.047***	0.056***	0.055***
	(0.012)	(0.010)	(0.008)	(0.008)
Backlogs of work	0.346***	0.291***	0.208***	0.203***
	(0.082)	(0.063)	(0.062)	(0.061)
Durable goods share of real consumption		0.896**		0.215
		(0.408)		(0.277)
Weighted average of car inflation rates			0.068***	0.064***
			(0.007)	(0.008)
Constant	−17.802***	−50.165***	−11.563***	−19.682*
	(4.123)	(15.901)	(3.252)	(10.375)
Observations	33	33	33	33
R^2	0.785	0.827	0.920	0.922
Adjusted R^2	0.771	0.809	0.912	0.911

Sources: Federal Reserve Bank of Cleveland; Harper Peterson and Co.; Baltic Exchange; IHS Markit Economics; Federal Reserve Bank of New York; Oxford Covid-19 Government Response Tracker; authors' calculations.

Note: Relative energy, food, and auto-related price inflation variables are created by subtracting median inflation from energy, food, and auto-related price inflation, respectively, and these are in monthly annualized terms. Backlogs of work variable is taken from IHS Markit Economics. Huber-White standard errors in parentheses. We do not report Newey-West standard errors because they can be unreliable in a sample as short as ours.

***$p < .01$, **$p < .05$, and *$p < .10$

services orders, which we interpret as a measure of supply chain disruptions (0.429); the share of goods in aggregate consumption, which captures the shift away from services during lockdowns (0.253); and auto price shocks, measured as a weighted average of auto-related inflation rates (new and used cars, car rentals, and car insurance) minus median inflation (0.191).

In multiple regressions, we find high explanatory power from a combination of three variables: energy price shocks, backlogs of work, and auto price shocks. A regression of headline shocks on these variables has an adjusted R^2 of 0.912. When all three are included, the goods share is not significant.

Figure 11 shows the actual and fitted values of headline shocks with the three key variables in the regression, along with the paths of the three variables. All three help explain the downward spike in headline inflation at the start of the pandemic, and they explain different parts of the subsequent high-inflation experience. For example, auto-related prices are important for the inflation run-up in summer 2021, the height of the chip shortage that impeded auto production. Both energy prices and backlogs help explain the 10 percentage point headline shock in March 2022. Energy prices explain the positive headline shock in June 2022 and the negative shocks from July to September.

In sum, we find that headline inflation shocks during the pandemic are well explained by some of the factors stressed in popular discussions of inflation.[16]

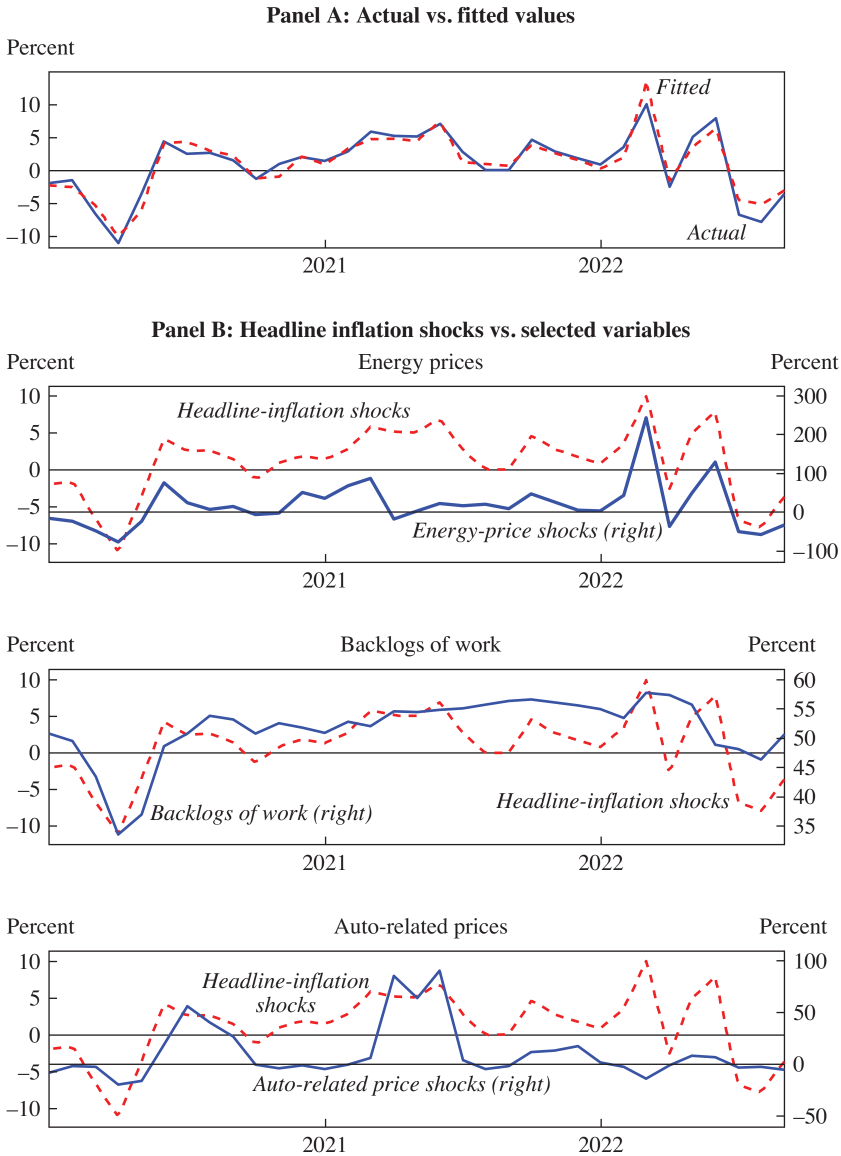
III.B. Accounting for the Rise in Inflation

Having analyzed both core inflation and deviations from core, we can do an accounting of the sources of the overall rise in inflation. We compare the twelve-month headline inflation rate in September 2022, 8.2 percent, to the rate of 1.3 percent in December 2020, when the early pandemic slump had pushed inflation down. We account for the 6.9 percentage point difference between these two inflation rates. Over the same period, twelve-month core (median) inflation increased 4.6 percentage points (from 2.3 percent to 7.0 percent).

In this exercise, we use the core inflation equation (column 4 of table 1) to determine the contributions to the rise in inflation of higher expected inflation, higher levels of *V/U*, and the pass-through variable *H*. We then use

16. The energy price and auto price variables also help explain headline shocks before the pandemic, but backlogs do not. Food price inflation is significant before the pandemic but not during the pandemic (see table 2A in the online appendix).

Figure 11. Explaining Headline Inflation Shocks



Sources: Authors' calculations; IHS Markit Economics.

Note: In panel A, headline inflation shocks denote the difference between headline and median CPI inflation. "Fitted" denotes fitted values of headline inflation shocks from the regression in table 2, column 3. In panel B, headline inflation shocks denote the difference between monthly annualized headline and median CPI inflation. Energy and auto-related price shocks variables are created by subtracting median inflation from energy and auto-related price inflation, respectively. These variables are in monthly annualized terms. Backlogs of work variable is taken from IHS Markit Economics.

our preferred equation for headline shocks (column 3 of table 2, panel B) to determine the shares of H to attribute to energy price shocks, backlogs, and auto price shocks. We use the same equation to account for the rise in the headline shock part of headline inflation. For each of the three contributors to headline shocks, we derive a total effect on the rise in headline inflation by summing the direct effect and the contribution to pass-through.[17]

Figure 12 shows the results. The combination of direct and pass-through effects of headline inflation shocks accounts for about 4.6 percentage points of the 6.9 percentage point rise in twelve-month inflation. Most of this 4.6 total reflects energy price shocks and backlogs of work, with total contributions of 2.7 and 1.7 percentage points, respectively. For each of these factors, roughly two-thirds of the contribution is the effect on current headline inflation and one-third is the pass-through into core. There is also a significant pass-through from past auto price shocks, reflecting the run-up in auto prices in summer 2021, but the direct effect on headline inflation has turned negative as these price increases have been partly reversed. A rise in expected inflation accounts for 0.5 percentage points.

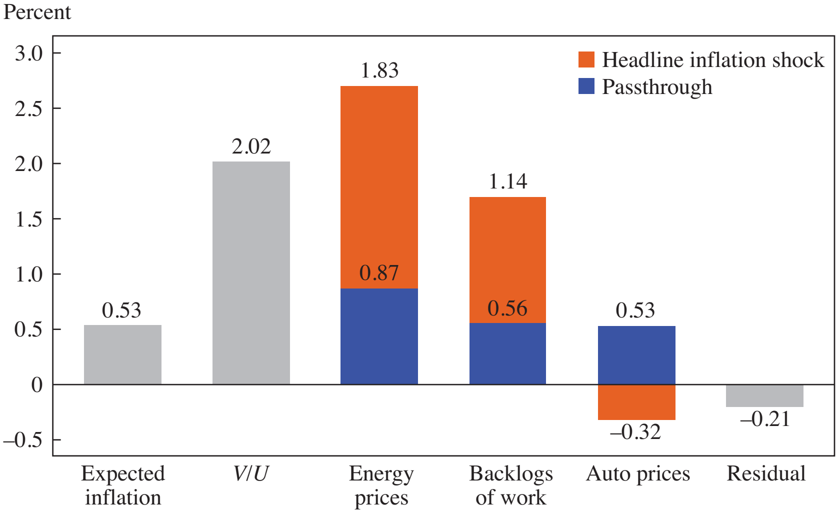
The contribution of *V/U* to the rise in twelve-month inflation is 2 percentage points, nearly a third of the total inflation increase. However, the rise in *V/U* explains more—nearly one-half—of the rise in *core* inflation, and as discussed above, the effect of *V/U* is rising over time. If we decompose the change in annualized one-month core inflation from December 2020 to September 2022 (a rise of 6.4 percentage points, from 1.9 percent to 8.3 percent), the contribution of *V/U* is 5 percentage points.

III.C. Why Has High Inflation Been Such a Surprise?

As inflation began to rise in March 2021, Federal Reserve chair Jerome Powell predicted that the increase would be "neither particularly large nor persistent" (Powell 2021a). At the Jackson Hole symposium that August,

17. The details of our calculations are as follows: (1) For a given month, we decompose core inflation into expected inflation, the effect of labor market tightness, the effect of past headline shocks ("pass-through effect"), and a residual (based on table 1, column 4). (2) Next, we decompose the pass-through effect into effects of energy shocks, backlogs, auto price shocks, and another residual by using their coefficients in our headline shock equation (table 2, panel B, column 3) and the twelve-month averages of the three variables. (3) Finally, we divide the current headline inflation shock into components due to the three variables, and another residual, using the same headline shock equation ("direct effects" of the variables). Having decomposed inflation in a given month, we subtract the average of each component over January–December 2020 from the average over October 2021–September 2022 to derive the decomposition of the twelve-month inflation change shown in figure 12. We report a single residual that combines the residuals from the different steps in our calculations.

Figure 12. Accounting for the Rise in Headline Inflation
(Decomposition of change in 12-month headline CPI inflation from December 2020 to September 2022; percentage points)



Sources: Authors' calculations; IHS Markit Economics.
Note: The total rise in twelve-month headline inflation is 6.94 percentage points (from 1.28 percent to 8.22 percent). The total rise in twelve-month core (median) CPI inflation over this period is 4.63 percentage points (from 2.34 percent to 6.98 percent). "Expected inflation" denotes contribution of change in long-term (SPF) inflation expectations to change in headline CPI inflation. *V/U* denotes contribution of change in ratio of vacancies to unemployed. "Energy prices" denotes contribution of relative energy prices. "Backlogs of work" denotes contribution of change in index from IHS Markit Economics. "Auto prices" denotes contribution of weighted average of auto-related prices. Based on estimates in table 1 (column 4) and table 2, panel B (column 3).

Powell remained sanguine, noting "the absence so far of broad-based inflation pressures" (Powell 2021b, 5). Powell's view was supported by the many economists on Krugman's (2021) "Team Transitory," including the authors of this paper (Spilimbergo and others 2021). Today, it is clear that inflation was much higher than we expected.

What accounts for these forecasting errors? One factor was unexpected: adverse shocks to headline inflation. These shocks include the unusual and persistent disruption of supply chains and the rise in energy prices associated with the war in Ukraine. On the other hand, part of the problem was flaws in our pre-pandemic understanding of inflation that recent experience has made apparent. There were three intertwined problems with conventional thinking.[18]

18. The analysis here overlaps with Furman (2022).

First, economists measured labor market tightness with the deviation of unemployment from its natural rate, typically as estimated by the CBO. As a result, they neglected the tightening of the labor market captured by the dramatic increase in the ratio of job vacancies to unemployed, although this was unexpected and did not occur until late 2021.[19]

Second, many (although not all) economists assumed that the effect of unemployment on inflation was linear and fairly small, based on estimates from the pre-pandemic era of stable inflation. As a result, even when they considered the possibility of an extreme tightening of the labor market, they expected the inflationary effects to be modest. Spilimbergo and others (2021), for example, predicted that if the unemployment rate fell to 1.5 percent, core inflation would rise only to 2.9 percent.

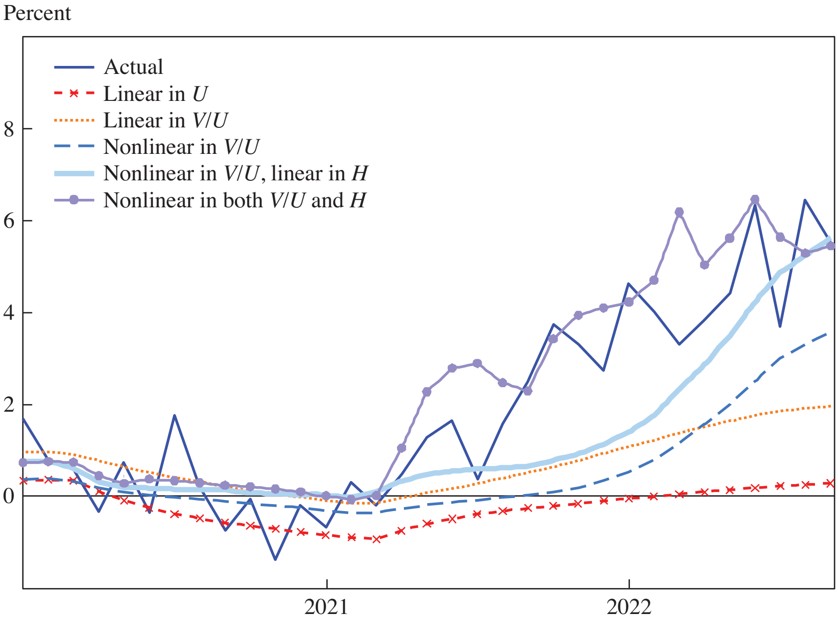
Finally, economists typically assumed explicitly or implicitly that deviations of headline inflation from core would not feed into core—they ignored the pass-through effect. If that effect had been accounted for, there would have been greater concern about core inflation in mid-2021, because at that point there had already been large headline inflation shocks, and prudent forecasters would have considered the risk of additional shocks as the economy reopened.

To illustrate these points, we compare the performance of alternative equations for monthly core inflation. We compare this paper's preferred equation to one that is linear in the twelve-month deviation of unemployment from the natural rate (as estimated by the CBO) and that excludes the pass-through variable H. This equation is similar to those estimated in much pre-pandemic work on the Phillips curve, including our own. To isolate the importance of different aspects of our specification, we change the traditional equation into our preferred one in steps: first replacing the unemployment measure of slack with V/U while maintaining a linear relation; then using a cubic rather than linear function of V/U; then adding H to the equation, first linearly and then as a cubic, which gives our preferred equation. We estimate each specification over the pre-pandemic era of 1985–2019 and then use the estimated equations to forecast core inflation during the pandemic.

Figure 13 shows the results. (The underlying regressions are in the online appendix.) We see again that our preferred core inflation equation performs well, as shown by the predicted path. We also see that the traditional equation with only a linear unemployment term performs quite poorly: it predicts a

19. In March 2021, the V/U ratio was 0.9, well below its pre-pandemic (January 2020) level of 1.2, with little indication that it would rise to above 2.0 by March 2022.

Figure 13. Predictions for Median CPI Inflation Gap during the Pandemic: Comparison across Models



Source: Authors' calculations.

Note: Figure reports predicted values based on monthly equations estimated for 1985–2019 (online appendix table 13A). Our preferred core inflation equation is shown by the predicted path in short dashes with circles. The predicted values from the traditional equation with only a linear unemployment (U) term is reported by dashes with crosses. The other fitted values in the figure show that each of our modifications to the traditional specification—the measure of slack, nonlinearity, and the pass-through variable (H)—contributes to the good fit of our final preferred equation.

decrease in inflation in 2020 and almost no increase since then, reflecting the fact that the twelve-month average of the unemployment rate has not fallen much below the CBO's natural rate (currently 4.4 percent). The other fitted values in figure 13 show that each of our modifications of the traditional specification—the measure of slack, nonlinearity, and the pass-through variable—contributes materially to the good fit of our final equation.[20]

Today we can see that, even before the pandemic, inflation equations fit the data better with tightness measured by a nonlinear function of V/U than

20. The online appendix presents the same comparison of specifications with core inflation measured by median PCE inflation. The results are similar to those for median CPI: the traditional equation fails to predict a significant rise in inflation; our preferred specification predicts most of the observed rise (although there is some underprediction since May 2022); and the measure of slack, nonlinearity, and the pass-through variable are all important.

with a linear function of *U*, and with a pass-through effect (see table 13A in the online appendix). Before 2020, however, the evidence on these points was not striking enough to influence the inflation models of most economists. Movements in *V/U* were strongly correlated with movements in unemployment, and we did not observe the extreme labor market tightness that has made nonlinearity obvious. Headline inflation shocks were smaller and less persistent than they have been since 2020, making the pass-through effect easy to miss.[21]

IV. Two Big Questions

We now move from explaining past inflation to considering the future. Like most economists, we presume that the Federal Reserve has the ability to rein in inflation if it raises interest rates by enough. What is less clear are the costs of doing so: Will containing inflation require a substantial slowing of the economy and increase in unemployment? Here we consider two factors that will help determine the answer: the behavior of the Beveridge curve, and the behavior of inflation expectations. There is considerable uncertainty about both issues.

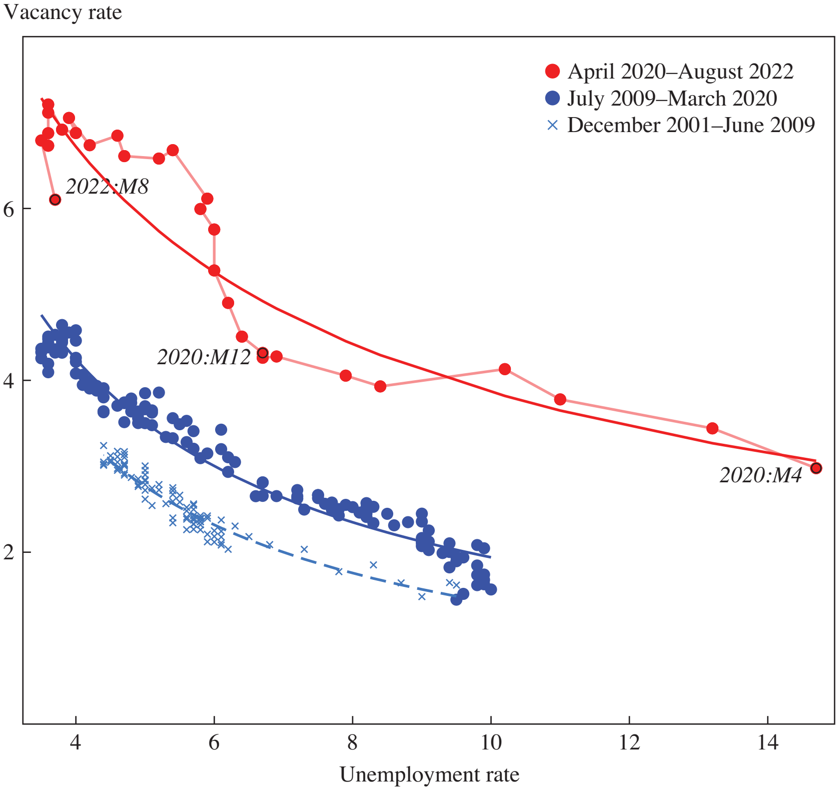
IV.A. The Beveridge Curve

The Beveridge curve is the relation between the unemployment rate and the vacancy rate. It is downward-sloping, reflecting the fact that a tightening of the labor market increases vacancies and reduces unemployment. As stressed by Blanchard, Domash, and Summers (2022), the Beveridge curve determines the relation between the unemployment rate and *V/U*, and therefore affects the level of unemployment needed to reduce inflation.

THE SHIFT IN THE CURVE Figure 14 plots the unemployment and vacancy rates from 2001 through August 2022. A stable Beveridge curve appears in different periods, but the curve has shifted at discrete points in time. The curve was stable from 2001 to 2009, then shifted outward and was stable again until March 2020. With the pandemic shutdown of April 2020, the curve abruptly shifted outward by a larger amount. Initially, the shift was a jump in the unemployment rate to 14.7 percent with little change in the vacancy rate; since then, the tightening of the labor market has moved the economy up the new Beveridge curve, and recent months have seen

21. Ball and Mazumder (2021) find a pass-through effect for the euro area but fail to find one for the United States. We can now see that the negative US result reflects an assumption that the effect is linear, which the data reject.

Figure 14. The Behavior of the Beveridge Curve, 2001–2022



Source: Bureau of Labor Statistics.

Note: December 2001 to June 2009 covers the Great Recession and the preceding expansion, based on National Bureau of Economic Research (NBER) business cycle dates. July 2009 to March 2020 covers the pre-COVID-19 expansion and the first month of the COVID-19 era. The figure reports log-linear curves fitted to each period. Rates are given as a percentage of the labor force.

unemployment rates close to pre-pandemic levels along with very high vacancy rates.[22]

Within a regime with a stable Beveridge curve, the curve is well approximated by a log-linear relationship between the unemployment and vacancy rates. Figure 14 shows log-linear curves that we estimate for the three periods since 2001.

22. The unemployment rate is U/(labor force), and we define the vacancy rate as V/(labor force), so the ratio of the two rates equals the V/U in our Phillips curve. Many researchers define the vacancy rate as V/(employment + V), but that distinction does not make a material difference for our analysis.

The outward shift in the Beveridge curve means that the labor market has become less efficient at matching unemployed workers with vacant jobs. It is not clear why that has happened, although recent work has suggested possible factors. Blanchard, Domash, and Summers (2022) cite increased reallocation of workers across firms, as captured by the gross level of hiring. Briggs (2022) cites decreased search intensity of unemployed workers, as indicated by a decline in the fraction who actively submit job applications.

Since we are not sure why the Beveridge curve has shifted, it is difficult to say whether temporary factors are responsible, in which case we should expect it to shift back at some point, or whether the shift is permanent. In August 2022, the last month shown in figure 14, *V* decreased noticeably with little change in *U*, but it is too soon to tell whether this is the start of a significant shift in the Beveridge curve. This issue is closely related to the debate between Blanchard, Domash, and Summers (2022) and the Federal Reserve's Figura and Waller (2022) about prospects for the labor market. Figura and Waller (2022) suggest that a cooling of demand can reduce the vacancy rate with little increase in unemployment, which will indeed be possible if the Beveridge curve shifts favorably. Blanchard, Domash, and Summers (2022) argue that this outcome is unlikely based on historical evidence. We will see that this issue is critical for the costs of reducing inflation.

THE RELATION BETWEEN UNEMPLOYMENT AND CORE INFLATION A log-linear Beveridge curve defines the vacancy rate v is a function of the unemployment rate u:

$$v = au^{b},\ a > 0,\ b < 0, \tag{2}$$

which implies a relation between the ratio *V/U* and the unemployment rate:

$$V/U = v/u = au^{b-1}. \tag{3}$$

If we substitute this expression for *V/U* in the Phillips curve, we obtain a relation between the core inflation gap (median inflation minus expected inflation) and the unemployment rate. This relation captures the unemployment-inflation trade-off facing policymakers as they stimulate or restrain demand and thereby move the economy along a stable Beveridge curve. In addition, this relation implies that there are now two possible shocks to the Phillips curve relationship: the Beveridge curve shock in addition to the more traditional cost-push shock.[23]

We derive this trade-off for two versions of the Beveridge curve: the ones estimated for the pre-pandemic period and the pandemic period (the solid lines in figure 14). In both cases, we use the monthly Phillips curve estimated for 1985–2022 (column 4 of table 1). In this exercise we set the headline shock variable H to zero.[24]

Figure 15 shows the results for the two Beveridge curves. A feature that jumps out in both cases is a striking nonlinearity: there is a sharp bend in the curve. At high unemployment rates, the relation is close to linear and flat. For example, for the pre-pandemic period, the slope is –0.28 at 8 percent unemployment and –0.31 at 6 percent, numbers that are roughly comparable to pre-pandemic estimates of the Phillips curve slope (Hazell and others 2022). However, the slope is –0.67 at 4 percent unemployment and rises dramatically to –2.8 at 3.5 percent unemployment.

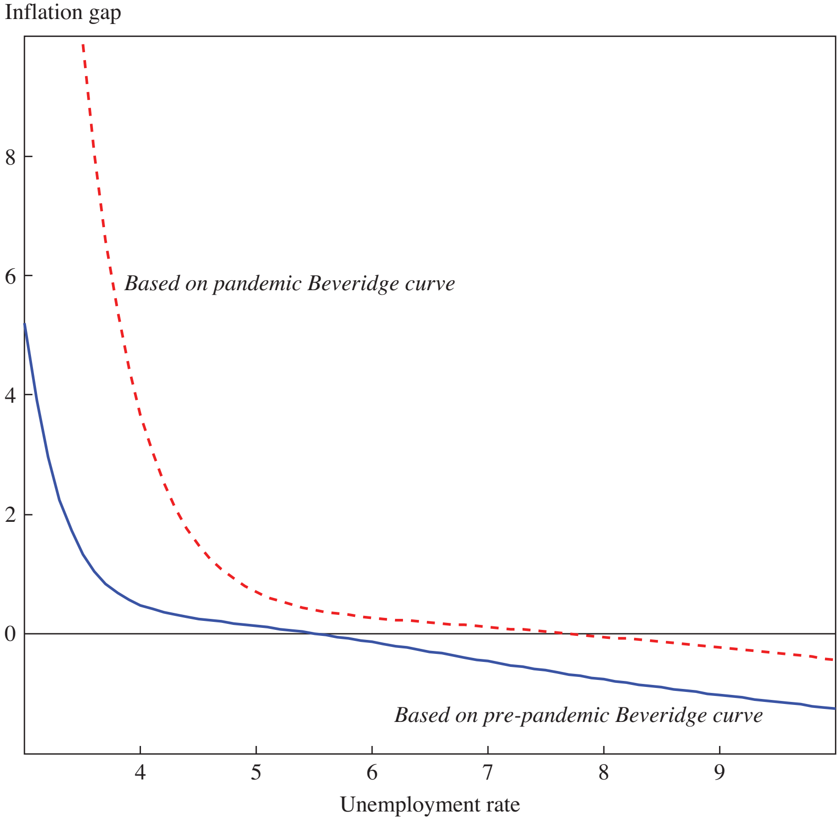
The shape of the curves in figure 15 supports Gagnon and Collins's (2019) view that the unemployment-inflation trade-off is steeper when unemployment is low. In our framework, this nonlinearity has two sources corresponding to the two relations from which the curves are derived. First, as seen in figure 6, V/U has a nonlinear effect on inflation, with a large marginal effect when V/U is high. Second, V/U is strongly nonlinear in U, with a large marginal effect when U is low. This second nonlinearity reflects the facts that both $1/U$ and V are convex in U, the latter because of the shape of the Beveridge curve.

The other message from figure 15 is that the unemployment-inflation trade-off has worsened during the pandemic: the inflation rate is now higher for any given unemployment rate, especially when unemployment is low. For example, at an unemployment rate of 4 percent, the core inflation gap is 0.5 percentage points with the old Beveridge curve and 3.7 percentage points with the pandemic Beveridge curve. This difference reflects the fact that 4 percent unemployment implies a much higher V/U with the pandemic curve. We will see that the shift in the unemployment-inflation trade-off,

23. These two shocks are not structural or independent. For instance, some shocks could increase production costs and simultaneously increase mismatch in the labor market. But they are also not identical: shocks to the Beveridge curve could be unrelated to cost-push shocks.

24. The estimated parameters in the Beveridge curves are $a = 13.9$ and $b = -0.85$ for the pre-pandemic (July 2009–March 2020) sample and $a = 15.5$ and $b = -0.60$ for the pandemic (April 2020–August 2022) sample. The latter period ends in August 2022 because the vacancy rate for September is not yet available.

Figure 15. Median CPI Inflation Gap versus Unemployment Rate for Different Beveridge Curves



Sources: Survey of Professional Forecasters; authors' calculations.
Note: "Inflation gap" denotes monthly annualized median CPI inflation minus long-term inflation expectations. Curves are derived from the estimates of the Phillips curve (table 1, column 4) and the Beveridge curves reported in the text.

if it persists, will make it costly for the Federal Reserve to reverse the pandemic-era rise in inflation.

THE NATURAL RATE OF UNEMPLOYMENT We can use the unemployment-inflation relationships in figure 15 to estimate the natural rate of unemployment and how it has changed during the pandemic. Following Friedman (1968), we define the natural rate as the unemployment rate at which actual inflation equals expected inflation. It is the unemployment rate that is sustainable in the long run.

One might think that the natural rate is the unemployment rate at which the inflation gap in figure 15—the difference between core inflation and long-term expected inflation—is zero. There is, however, a subtle complication: core inflation is *median* inflation but expected inflation is a survey measure of expected *headline* inflation, which could differ slightly from expected median inflation. Over 1985–2019, median inflation exceeded headline inflation by an average of about 0.2 percentage points (which means on average there was a slight left skewness in the distribution of industry inflation rates). We therefore assume that long-term expected core inflation is 0.2 percentage points higher than expected headline inflation. This assumption implies that the natural rate of unemployment is the rate at which the inflation gap in figure 15 is 0.2.

Based on this definition, the natural rate of unemployment is 4.8 percent for the unemployment-inflation relation derived from the pre-pandemic Beveridge curve in figure 15 and 6.5 percent for the pandemic-era Beveridge curve. The 4.8 estimate is close to other natural rate estimates for the pre-pandemic era (for example, the CBO's natural rate averaged 5.2 percent over 1985–2019). Our finding that the natural rate has risen 1.7 percentage points during the pandemic is roughly consistent with Crump and others (2022) and Blanchard, Domash, and Summers (2022), who report natural rate increases of 2.0 and 1.3 percentage points, respectively. In our framework, the increase has resulted from the outward shift in the Beveridge curve, and the natural rate will fall if the curve shifts back toward its pre-pandemic position.

We should emphasize that estimates of the natural rate of unemployment are imprecise. This is true both in general (Staiger, Stock, and Watson 1997) and in particular for our calculations because they depend on our calibration of the difference between expected median and expected headline inflation. Small changes in that number imply substantial changes in our natural rate estimates. That said, the result that the outward shift in the Beveridge curve has increased the natural rate is robust.[25]

IV.B. Will Inflation Expectations Remain Anchored?

In the two decades before the pandemic, long-term inflation expectations were well-anchored at the Federal Reserve's inflation target, and

25. If we assume that the difference between expected median and expected headline inflation is zero, then the estimated natural rates for the pre-pandemic and pandemic periods are 5.5 percent and 7.8 percent, respectively. If we assume that the difference in expected inflation is 0.4 percentage points (which is the average difference between median and headline inflation in the decade before 2020), the estimated natural rates are 4.0 percent and 5.3 percent.

this anchoring made it easier to return actual inflation to target when it was pushed away temporarily. Looking forward, if expectations remain anchored, then inflation will again return to target once the labor market normalizes and the economy moves beyond the unusual shocks of the pandemic.

However, the anchoring of inflation expectations is not immutable. Anchoring has occurred because the Federal Reserve has built a track record of reversing short-run movements in inflation and returning inflation to target. Presumably a large enough and persistent enough rise in inflation would eventually lead people to revise their expectations upward, which in turn would push actual inflation even higher. That outcome would worsen the unemployment-inflation trade-off and increase the costs of reining in inflation.

There are hints that a de-anchoring of expectations may already have begun. As shown in figure 2 above, ten-year expected inflation in the SPF has risen from 2.2 percent in 2019:Q4 to 2.8 percent in 2022:Q3. Five-year expectations in the University of Michigan Survey of Consumers have risen from 2.2 percent in December 2019 to 2.7 percent in September 2022.[26]

Will these modest increases in expected inflation be reversed as the Federal Reserve takes action to control inflation? Or are we seeing the beginning of a substantial de-anchoring? It is hard to know, but we seek to inform discussions of the issue by carefully examining the response of expectations to inflation movements, both in the pandemic period and earlier.[27]

A SIMPLE MODEL OF INFLATION EXPECTATIONS We posit a simple equation in which expectations evolve in response to movements in headline inflation:

$$\pi_t^e = \gamma\pi_{t-1}^e + (1-\gamma)\pi_t, \tag{4}$$

where π^e is expected inflation and π is actual headline inflation. The parameter γ captures the degree of anchoring. For $\gamma = 1$, expected inflation is constant regardless of actual inflation behavior. For $\gamma = 0$, expected inflation adjusts one-for-one with current inflation.

26. Surveys of Consumers, University of Michigan, "Times Series Data," table 33: Expected Change in Prices during the Next 5 Years, https://data.sca.isr.umich.edu/data-archive/mine.php.

27. See also Reis (2021) and Hilscher, Raviv, and Reis (2022), who examine the distribution of expectations across individual survey respondents to assess the risk of de-anchoring.

We consider the evolution of expected inflation over some period starting at $t = \tau$. By repeatedly substituting the equation for π^e into itself, we obtain

$$\pi_t^e = (1-\gamma)\sum_{i=0}^{t-\tau-1}\gamma^i\pi_{t-i} + \gamma^{t-\tau}\pi_\tau^e, t > \tau. \tag{5}$$

We estimate γ with the SPF's quarterly series for ten-year expected inflation. We account for the fact that the current quarter's inflation rate is not known when a ten-year forecast is made by replacing π_t (the first term in the summation) with the current-period (SPF) expectation of π_t, a now-cast that is reported at the same time. We denote this expectation by ${}_t\pi_t$. We also add an error term to the equation to capture other influences on expectations, yielding:

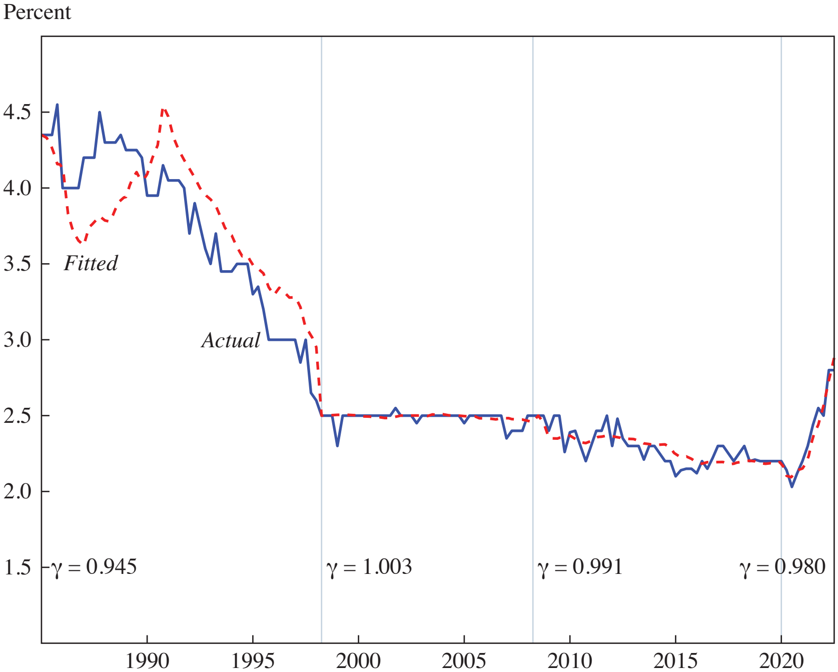
$$\pi_t^e = (1-\gamma)\,{}_t\pi_t + (1-\gamma)\sum_{i=1}^{t-\tau-1}\gamma^i\pi_{t-i} + \gamma^{t-\tau}\pi_\tau^e + \epsilon_t, t > \tau. \tag{6}$$

We estimate γ, the single parameter in this equation, with nonlinear least squares.

ANCHORING IN SEVERAL ERAS We examine the behavior of expectations in several time periods. Specifically, we divide the data from 1985 to the present into four periods for which we have reason to believe that expectations behaved differently. Figure 16 shows the path of expected inflation since 1985, the estimated γ for each period, and the associated fitted values for expected inflation. Our results and interpretation for the four periods are as follows:

- 1985:Q1–1998:Q1: this is the period before anchoring, when the actual CPI inflation rate drifted down from about 4 percent to 2.5 percent and expectations followed. The estimated γ is 0.945, the lowest for the four periods.
- 1998:Q2–2008:Q2: the start of this period is the beginning of the anchoring regime identified by Ball and Mazumder (2018). Actual inflation fluctuated but expected inflation was almost constant at 2.5 percent, and the estimated γ is 1.003.
- 2008:Q3–2019:Q4: this is the period following the Great Recession, when inflation repeatedly fell short of the Federal Reserve's target, albeit by small amounts. It appears that this experience produced some de-anchoring, with expected inflation falling. The estimated γ is 0.991.
- 2020:Q1–2022:Q3: the pandemic period in which expected inflation has risen somewhat. The estimated γ is 0.980, suggesting that anchoring has become weaker than it was before the pandemic.

Figure 16. Actual and Fitted Inflation Expectations



Sources: Survey of Professional Forecasters; authors' calculations.
Note: Figure reports actual values of long-term CPI inflation expectations and fitted values for several periods from the partial-adjustment model described in the text. The parameter γ indicates the degree of anchoring of inflation expectations in each period.

In what follows, we use these historical experiences as guides to what might happen in the future.

V. Scenarios for Future Inflation

Where is inflation heading? We will not offer unconditional forecasts. The path of inflation will depend on how quickly the Federal Reserve raises interest rates and how those actions and other factors affect the labor market. We will leave forecasts concerning those issues to others, and forecast inflation paths conditional on paths for unemployment. This exercise will help us see how much the Federal Reserve needs to raise unemployment to return inflation to an acceptable level.

One unemployment path we consider is the one projected by members of the Federal Open Market Committee (FOMC) in their most recent

(September 2022) *SEP*. In this scenario, the unemployment rate rises only modestly from its current level, peaking at 4.4 percent at the end of 2023. We also consider a more pessimistic forecast from the International Monetary Fund (IMF)'s October 2022 *World Economic Outlook* in which (in the quarterly data underlying the report) unemployment peaks at 5.6 percent in 2024, and a much more pessimistic scenario suggested by Summers (Mellor 2022) in which unemployment rises to 7.5 percent for two years. Summers suggests that unemployment must rise that much to return inflation to the Federal Reserve's target.

Once we assume a path for the unemployment rate, there is still uncertainty about the path of inflation because it will depend on the behavior of the Beveridge curve and of expectations. We construct forecasts for both optimistic and pessimistic assumptions about these factors.

In all our simulations, we set headline inflation shocks to zero starting in October 2022. This is a natural benchmark because historically headline shocks have been unpredictable and not persistent.[28] However, it is important to keep in mind that the future could bring either positive or negative headline shocks. We might see inflationary shocks resulting from a worsening of the war in Ukraine or new disruptions of production as the pandemic waxes and wanes. We might see disinflationary shocks if energy prices fall or other supply factors improve. (Currently, oil futures curves suggest that crude oil prices are expected to decrease in coming years.) Either way, there could be major movements in inflation that are unrelated to monetary policy.

V.A. Alternative Assumptions about the Beveridge Curve and Expectations

We consider the following scenarios.

THE BEVERIDGE CURVE Our pessimistic case for the Beveridge curve is that it remains in its position during the pandemic to date, as captured by the log-linear relation we have estimated (see figure 14). This means that the factors that have worsened the ability of the labor market to match workers to jobs, whatever they are, persist.

Our other scenario is that the Beveridge curve shifts back quickly to its pre-pandemic position (see figure 14). Specifically, starting from the pandemic era curve in September, the curve shifts one-quarter of the way

28. The serial correlation of headline inflation shocks is low: an AR(1) specification for the monthly headline inflation shock yields an estimated coefficient of 0.4 for 1985–2019 and 0.5 for 2020–2022.

toward its pre-pandemic position every month, which means the outward shift during the pandemic is almost entirely reversed after about six months.[29]

EXPECTATIONS We specify paths for expected inflation at the monthly frequency. In all cases we start with expected inflation in September 2022 at 2.8 percent, the level reported in the SPF for 2022:Q3. We consider three scenarios for the evolution of expectations starting in October.

In our most optimistic scenario, confidence in the Federal Reserve's commitment to low inflation ensures that expected inflation quickly reverts to its pre-pandemic level of 2.2 percent. Specifically, it moves one-quarter of the way each month.

A second scenario is that expected inflation continues to respond to actual inflation as our estimates suggest it has so far during the pandemic. That is, expected inflation follows the pandemic era process: $\pi_t^e = \gamma\pi_{t-1}^e + (1 - \gamma)\pi_t$ with $\gamma = 0.980$ at the quarterly frequency. We set γ equal to the cube root of 0.980 in our monthly simulations.

Finally, we consider a variation on the second scenario with $\gamma = 0.944$ at the quarterly frequency, which is the estimated anchoring parameter for the 1985–1998 period. We view this case as quite pessimistic: expectations behave as they did before 1998, which means that all of the progress in anchoring expectations since then is lost.

V.B. Deriving Inflation Paths

For given assumptions about the Beveridge curve and inflation expectations and a given path of the unemployment rate, and starting from actual data through September 2022, we construct a monthly simulation of the economy. For each month, the steps are:

- Use the Beveridge curve to derive *V/U* given the assumed *U*, and compute the twelve-month average of *V/U*.
- Compute the twelve-month headline shock *H* given zero monthly shocks starting in October 2022 and the actual shocks before that. The twelve-month average declines to zero in September 2023.
- Given the twelve-month *V/U* and *H*, compute the core inflation gap from the monthly Phillips curve (column 4 of table 1).
- Given the core inflation gap and the level of expected inflation in the previous month, derive the current levels of core inflation and

29. If $v^*(u)$ and $v^{**}(u)$ are the pre-pandemic and pandemic Beveridge curves, then the curve in October 2022 is $.75v^{**}(u) + .25v^*(u)$. After October 2022, the curve in month t is $v_t(u) = .75v_{t-1}(u) + .25v^*(u)$.

expected inflation from the equation for expected inflation (except in the most optimistic expectations scenario, in which expected inflation moves one-quarter of the way toward 2.2 percent).[30]

These steps yield a monthly series for core inflation starting in October 2022. By the assumption that future headline shocks are zero, the monthly path of headline inflation is the same. We aggregate over time to derive a twelve-month path of core inflation. The twelve-month path of headline inflation converges to core in September 2023.

V.C. Inflation Paths for the FOMC's Unemployment Forecasts

In considering possible paths for unemployment, a natural starting point is the forecasts of Federal Reserve policymakers, which are reported in the *SEP*s released after every other FOMC meeting. The most recent *SEP* as this paper is written is the one for September 21, 2022. In these forecasts, the unemployment rate rises only modestly over time and peaks in late 2023 at 4.4 percent. This unemployment rate is low by historical standards and equals the CBO's current estimate of the natural rate. According to the *SEP*, the economy will experience low unemployment at the same time as inflation falls back to the Federal Reserve's target.

The *SEP* forecasts the unemployment rate in the fourth quarters of 2022, 2023, and 2024. We construct a monthly unemployment path by assigning each fourth quarter forecast to November and then interpolating, starting with the actual unemployment rate of 3.5 percent in September 2022.

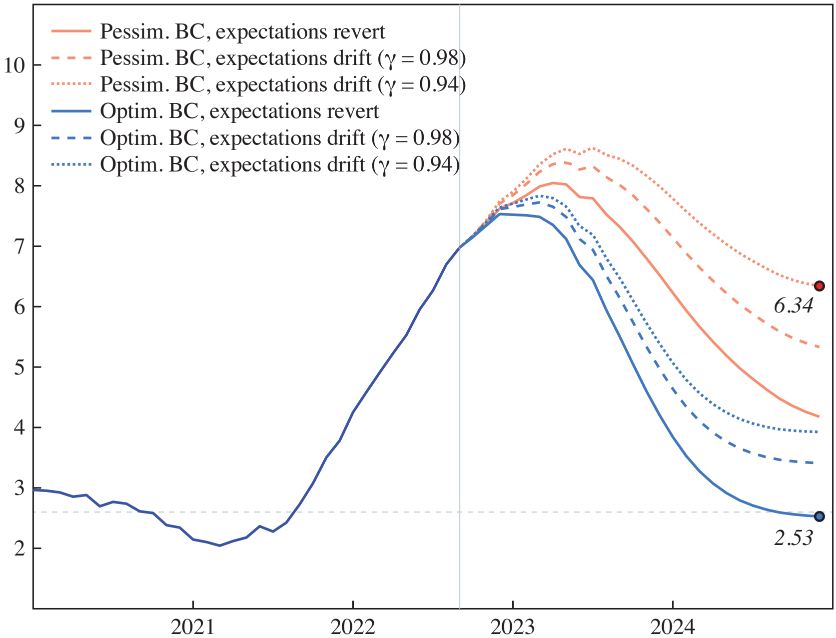
Figure 17 shows simulated paths of twelve-month core (median CPI) inflation for the *SEP* unemployment path and our different Beveridge curve and expectations scenarios. Online appendix figure 17A repeats this exercise for median PCE inflation, yielding similar results. To illustrate the mechanisms behind the results, figure 18 shows the paths of all simulated variables for one case, the pessimistic Beveridge curve and intermediate expectations assumption.

The different core inflation paths in figure 17 have some common features. They all rise from the current level of 7 percent and peak at some point between December 2022 and July 2023, reflecting the fact that the twelve-month average of *V/U* continues to rise even as somewhat higher unemployment reduces the current *V/U*. Eventually core inflation starts to decline as *V/U* continues to fall and the pass-through effects of past headline shocks die out.

30. Except in the most optimistic scenario, we use the equations $\pi_t = \pi_t^e$ + core gap and $\pi_t^e = \gamma\pi_{t-1}^e + (1-\gamma)\,\pi_t$. Given the core gap and π_{t-1}^e, we can solve the two equations for π_t and π_t^e.

Figure 17. Scenarios for Core CPI Inflation Conditional on September 2022 FOMC Unemployment Forecasts

Percent, 12-month



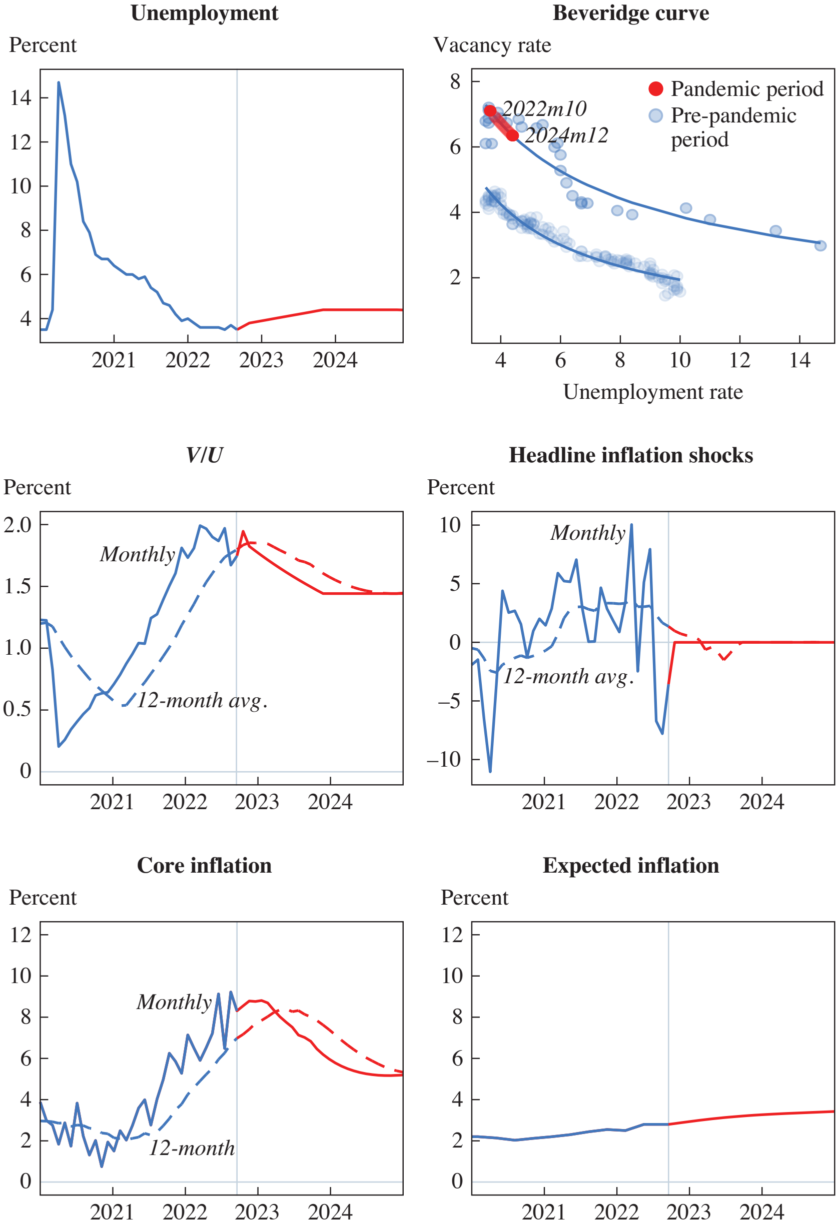
Sources: Authors' calculations; Federal Reserve Bank of Atlanta.

Note: Unemployment forecast from the *Summary of Economic Projections* of the Federal Open Market Committee (FOMC 2022), published in September 2022, which provides numbers for the fourth quarters of 2022, 2023, 2024, and 2025. We assign those forecasts to November of each year and interpolate a monthly unemployment series starting from the actual value of 3.5 percent in September 2022. The vertical line indicates September 2022. Core inflation denotes CPI median inflation. The horizontal dashed line shows the 2.6 percent target for median CPI based on the 2 percent PCE target as reported on the Federal Reserve Bank of Atlanta's Underlying Inflation Dashboard.

The levels of inflation, however, vary greatly across the different scenarios. With the most optimistic assumptions about both the Beveridge curve and expected inflation, core inflation peaks at 7.5 percent and falls to 2.5 percent in December 2024. With the most pessimistic assumptions, core inflation peaks at 8.6 percent and its December 2024 level is 6.3 percent, only 0.7 percentage points below the current level.

While both the Beveridge curve and inflation expectations affect the inflation path, the former is more important. If the Beveridge curve shifts back to its pre-pandemic position, the December 2024 inflation rate ranges from 2.5 to 3.9 percent depending on the expectations scenario. In contrast, if the

Figure 18. Scenario Conditional on September 2022 FOMC Unemployment Forecasts with Pessimistic Beveridge Curve and Drifting Expectations



Source: Authors' calculations.

Note: Figure reports scenario with COVID-19-era Beveridge curve and drifting expectations ($\gamma = 0.98$). Observations up to September 2022 are shown to the left of the vertical line; projections thereafter to the right. Core inflation denotes CPI median inflation.

Beveridge curve does not shift back, the inflation rate always stays above 4 percent. With the pandemic era Beveridge curve, a peak unemployment rate of 4.4 percent is not high enough to reduce *V/U* to a noninflationary level.

In interpreting these results, one nuance is that we forecast core inflation as measured by the weighted median CPI, whereas the Federal Reserve targets a 2 percent inflation rate in the PCE deflator. According to the Federal Reserve Bank of Atlanta's Underlying Inflation Dashboard, the Federal Reserve's target is equivalent to a 2.6 percent target for median CPI inflation, given the historical difference between the average levels of median CPI and headline PCE inflation. The upshot is that our most optimistic forecast for December 2024, a core inflation rate of 2.53 percent, is slightly below the Federal Reserve's target. In all the other scenarios, however, inflation stays above the target.

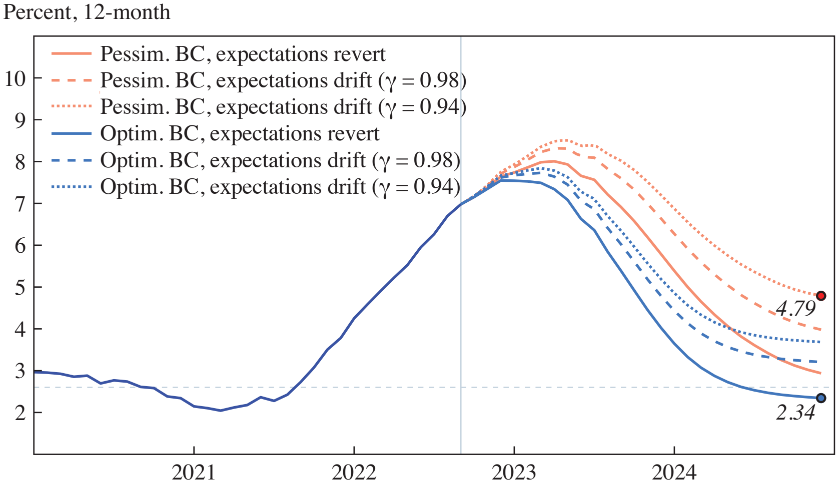
V.D. Inflation Paths with Higher Unemployment

If the *SEP*'s unemployment path risks leaving inflation at a high level, how much higher must unemployment rise to more reliably meet the Federal Reserve's inflation goal? To shed light on this question, we consider two other unemployment paths. One is based on unemployment forecasts for the United States in the IMF's *World Economic Outlook* of October 2022. These forecasts are more pessimistic than the Federal Reserve's: the unemployment rate rises to 5.6 percent in the second half of 2024. We construct a monthly unemployment scenario by assigning the IMF's quarterly forecasts to the middle month of each quarter and interpolating. The other path is based on Summers's highly pessimistic suggestion that reversing the rise in inflation will require two years of 7.5 percent unemployment (Mellor 2022). In this scenario, we assume that the unemployment rate rises linearly from its September 2022 level to 7.5 percent in January 2023 and then stays at 7.5 percent through December 2024.

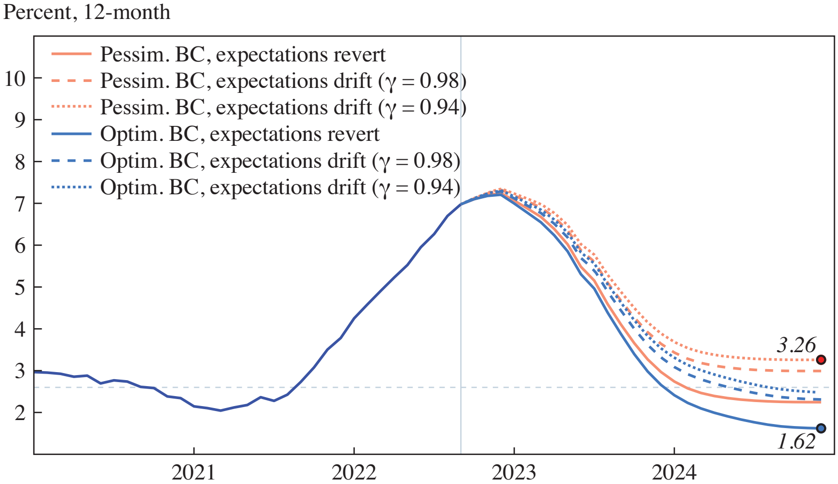
Figure 19, panel A, shows the inflation paths conditional on the IMF unemployment forecasts and our alternative Beveridge curve and expectations assumptions. As one would expect, higher unemployment lowers inflation: the December 2024 inflation level ranges from 2.3 to 4.8 percent, compared to 2.5 to 6.3 percent for the *SEP* unemployment path. Yet inflation still levels off above the Federal Reserve's target in most cases. Here, the behavior of inflation expectations is critical. Even with the more pessimistic Beveridge curve, median CPI inflation falls to 2.9 percent, only a bit above the implicit 2.6 percent target, if expected inflation reverts to its pre-pandemic level.

Figure 19. Scenarios for Core CPI Inflation Conditional on Alternative Unemployment Paths

Panel A: Conditional on IMF staff unemployment path



Panel B: Conditional on higher unemployment path



Sources: Authors' calculations; Federal Reserve Bank of Atlanta.

Note: Vertical line indicates September 2022. Core inflation denotes CPI median inflation. Panel A based on the IMF staff forecast for the quarterly path of unemployment underlying the October 2022 IMF *World Economic Outlook* report. Quarterly forecasts are allocated to the second month of each quarter, and a monthly path is obtained via interpolation. Panel B based on a higher unemployment path that assumes 7.5 percent unemployment during 2023 and 2024 as suggested by Summers (Mellor 2022). In this scenario, the unemployment rate rises linearly from its September 2022 level to 7.5 percent in January 2023 and remains at that level through December 2024. Horizontal dashes show 2.6 percent target for median CPI based on 2 percent PCE target reported on the Federal Reserve Bank of Atlanta's Underlying Inflation Dashboard.

Figure 19, panel B, shows the inflation paths for the scenario with two years of 7.5 percent unemployment. In this case, the differences across Beveridge curve and expectations assumptions are relatively small. The December 2024 inflation rates are clustered around 2.6 percent, with each less than 1 percentage point away from that level. Our analysis suggests, therefore, that this scenario's unemployment path robustly brings inflation close to the Federal Reserve's goal. Unfortunately, the cost is a painful and prolonged increase in unemployment. Moreover, a comparison of all the scenarios reported in figures 17–19 reveals that the sacrifice ratio, defined here as the additional unemployment required to reduce inflation by an extra percentage point by December 2024, is always larger for a greater reduction in inflation from the level today.

VI. Conclusion

Yogi Berra observed that "it's tough to make predictions, especially about the future." This aphorism applies to the study of US inflation.

Looking backward, we can account fairly well for inflation behavior during the pandemic. A tight labor market has pushed up core inflation, headline inflation has deviated from core because of sharp rises in energy and auto prices and supply chain problems, and pass-through from these headline shocks has magnified the rise in core. All of these factors have been prominent in recent discussions of inflation. We contribute a simple framework in which we quantify their roles. We find that the combination of direct and pass-through effects from headline inflation shocks accounts for about 4.6 percentage points of the 6.9 percentage point rise in twelve-month inflation between the end of 2020 and September 2022. A rise in expected inflation accounts for 0.5 percentage points, and the rise in labor market tightness (measured by the ratio of vacancies to unemployment) accounts for 2 percentage points. The role of labor market tightness is rising over time.

Looking forward, we can forecast inflation if we specify the path of unemployment and the future behavior of the Beveridge curve and inflation expectations. There is much uncertainty about these factors, so it is difficult to make unconditional predictions. Yet we have one broad finding: the forecasts of Federal Reserve policymakers—that inflation will return to target while unemployment rises only to 4.4 percent—are reasonable only under quite optimistic assumptions about both the Beveridge curve and expectations. If the behavior of either proves less benign, then reducing inflation is likely to require higher unemployment than the Federal Reserve anticipates.

While our simple framework explains recent inflation fairly well, future research might improve it along many dimensions. For example, researchers should continue to refine the measurement of core inflation, of labor market tightness, and of inflation expectations. We should try to better understand the nonlinear effects of tightness and past headline shocks on core inflation. We also need more work on why the Beveridge curve shifts and why inflation expectations become anchored or de-anchored.

ACKNOWLEDGMENTS We are grateful to Philip Barrett, Olivier Blanchard, Marijn Bolhuis, Nigel Chalk, Judd Cramer, Benjamin Friedman, Jason Furman, Pierre-Olivier Gourinchas, N. Gregory Mankiw, Wilson Powell, Ayşegül Şahin, Antonio Spilimbergo, and James Stock for valuable comments. We thank Regis Barnichon for providing us with historical job openings data; Mattia Coppo, Yiyang Han, Nino Kodua, and Kyung Woong Koh for excellent research support; and Eduard Laurito and Gabi Ionescu for administrative support. All errors remaining are our own. The views expressed here are those of the authors and do not necessarily represent those of the IMF, its Management and Executive Board, or IMF policy.

References

Abraham, Katharine, John Haltiwanger, and Lea Rendell. 2020. "How Tight Is the US Labor Market?" *Brookings Papers on Economic Activity*, Spring, 97–138.

Ball, Laurence, Daniel Leigh, Prachi Mishra, and Antonio Spilimbergo. 2021. "Measuring U.S. Core Inflation: The Stress Test of COVID-19." Working Paper 29609. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29609.

Ball, Laurence, and N. Gregory Mankiw. 1994. "Asymmetric Price Adjustment and Economic Fluctuations." *Economic Journal* 104, no. 423: 247–61.

Ball, Laurence, and Sandeep Mazumder. 2011. "Inflation Dynamics and the Great Recession." *Brookings Papers on Economic Activity*, Spring, 337–81.

Ball, Laurence, and Sandeep Mazumder. 2018. "A Phillips Curve with Anchored Expectations and Short-Term Unemployment." *Journal of Money, Credit and Banking* 51, no. 1: 111–37.

Ball, Laurence, and Sandeep Mazumder. 2021. "A Phillips Curve for the Euro Area." *International Finance* 24, no. 1: 2–17.

Barnichon, Regis. 2010. "Building a Composite Help-Wanted Index." *Economics Letters* 109, no. 3: 175–78.

Barnichon, Regis, Luiz E. Oliveira, and Adam H. Shapiro. 2021. "Is the American Rescue Plan Taking Us Back to the '60s?" *FRBSF Economic Letter*, October 18. Federal Reserve Bank of San Francisco.

Barnichon, Regis, and Adam Hale Shapiro. 2022. "What's the Best Measure of Economic Slack?" *FRBSF Economic Letter*, February 22. Federal Reserve Bank of San Francisco.

Bernanke, Ben S. 2004. "The Great Moderation." Remarks made at the meetings of the Eastern Economic Association, Washington, DC, February 20.

Blanchard, Olivier. 2021. "In Defense of Concerns over the $1.9 Trillion Relief Plan." Blog post, Realtime Economics, Peterson Institute for International Economics, February 18. https://www.piie.com/blogs/realtime-economic-issues-watch/defense-concerns-over-19-trillion-relief-plan.

Blanchard, Olivier. 2022. "Why I Worry about Inflation, Interest Rates, and Unemployment." Blog post, Realtime Economics, Peterson Institute for International Economics, March 14. https://www.piie.com/blogs/realtime-economic-issues-watch/why-i-worry-about-inflation-interest-rates-and-unemployment.

Blanchard, Olivier, Alex Domash, and Lawrence H. Summers. 2022. "Bad News for the Fed from the Beveridge Space." Policy brief. Washington: Peterson Institute for International Economics. https://www.piie.com/publications/policy-briefs/bad-news-fed-beveridge-space.

Bolhuis, Marijn A., Judd N. L. Cramer, and Lawrence H. Summers. 2022. "Comparing Past and Present Inflation." Working Paper 30116. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30116.

Borenstein, Severin, A. Colin Cameron, and Richard Gilbert. 1997. "Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes?" *Quarterly Journal of Economics* 112, no. 1: 305–39.

Briggs, Joseph. 2022. “The Beveridge Curve Debate: Has Match Efficiency Really Declined?” Goldman Sachs Economics Research, August 7, 2022.

Bryan, Michael F., and Stephen G. Cecchetti. 1994. “Measuring Core Inflation.” In *Monetary Policy*, edited by N. Gregory Mankiw. Chicago: University of Chicago Press.

Crump, Richard K., Stefano Eusepi, Marc Giannoni, and Ayşegül Şahin. 2022. “The Unemployment-Inflation Trade-off Revisited: The Phillips Curve in COVID Times.” Working Paper 29785. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29785.

di Giovanni, Julian, Şebnem Kalemli-Özcan, Alvaro Silva, and Muhammed A. Yildirim. 2022. “Global Supply Chain Pressures, International Trade, and Inflation.” Working Paper 30240. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30240.

Dolmas, Jim. 2005. “Trimmed Mean PCE Inflation.” Working Paper 0506. Federal Reserve Bank of Dallas Research Department. https://www.dallasfed.org/~/media/documents/research/papers/2005/wp0506.pdf.

Dolmas, Jim, and Evan F. Koenig. 2019. “Two Measures of Core Inflation: A Comparison.” Federal Reserve Bank of St. Louis *Review* 101, no. 4: 245–58.

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

European Central Bank. 2014. “Indirect Effects of Oil Price Developments on Euro Area Inflation.” *Monthly Bulletin*, December 2014, 54–56.

Faberman, R. Jason, Andreas I. Mueller, Ayşegül Şahin, and Giorgio Topa. 2020. “The Shadow Margins of Labor Market Slack.” *Journal of Money, Credit and Banking* 52, no. S2: 355–91.

Federal Open Market Committee (FOMC). 2022. *Summary of Economic Projections*. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/monetarypolicy/files/fomcprojtabl20220921.pdf.

Figura, Andrew, and Chris Waller. 2022, “What Does the Beveridge Curve Tell Us about the Likelihood of a Soft Landing?” FEDS Notes. Washington: Board of Governors of the Federal Reserve System.

Friedman, Milton. 1968. “The Role of Monetary Policy.” *American Economic Review* 58, no. 1: 1–17.

Furman, Jason. 2022. “America’s Wage-Price Persistence Must Be Stopped.” *Project Syndicate*, August 2. https://www.project-syndicate.org/commentary/inflation-wage-price-spiral-requires-fed-tightening-by-jason-furman-2022-08.

Furman, Jason, and Wilson Powell III. 2021. “What Is the Best Measure of Labor Market Tightness?” Blog post, Realtime Economics, Peterson Institute for International Economics, November 22. https://www.piie.com/blogs/realtime-economic-issues-watch/what-best-measure-labor-market-tightness.

Gagnon, Joseph E., and Christopher G. Collins. 2019. “Low Inflation Bends the Phillips Curve.” Working Paper 19-6. Washington: Peterson Institute for International Economics. https://www.piie.com/publications/working-papers/low-inflation-bends-phillips-curve.

Gordon, Robert J. 1975. “Alternative Responses of Policy to External Supply Shocks.” *Brookings Papers on Economic Activity*, no. 1: 183–206.

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.

Hilscher, Jens, Alon Raviv, and Ricardo Reis. 2022. “How Likely Is an Inflation Disaster?” Working Paper. Social Science Research Network, April 27. https://ssrn.com/abstract=4083404.

International Monetary Fund (IMF). 2022. *World Economic Outlooks: Countering the Cost-of-Living Crisis*. Washington: Author. https://www.imf.org/en/Publications/WEO/Issues/2022/10/11/world-economic-outlook-october-2022.

Krugman, Paul. 2021. “Wonking Out: I’m Still on Team Transitory.” *New York Times*, September 10. https://www.nytimes.com/2021/09/10/opinion/transitory-inflation-covid-consumer-prices.html.

Medoff, James L., and Katherine G. Abraham. 1982. “Unemployment, Unsatisfied Demand for Labor, and Compensation Growth in the United States, 1956–1980.” In *Workers, Jobs, and Inflation*, edited by Martin Baily. Washington: Brookings Institution.

Mellor, Sophie. 2022. “5 Years at 6% Unemployment or 1 year at 10%: That’s What Larry Summers Says We’ll Need to Defeat Inflation.” *Fortune Magazine*, June 21.

Mortensen, Dale T., and Christopher Pissarides. 1999. “Job Reallocation, Employment Fluctuations and Unemployment.” In *Handbook of Macroeconomics, Volume 1A*, edited by John B. Taylor and Michael Woodford. Amsterdam: North Holland.

Owyang, Michael T., and E. Katarina Vermann. 2014. “Rockets and Feathers: Why Don’t Gasoline Prices Always Move in Sync with Oil Prices?” *Regional Economist*, Federal Reserve Bank of St. Louis, October 1.

Phillips, A. W. 1958. “The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957.” *Economica* 25, no. 100: 283–99.

Powell, Jerome. 2021a. “Oversight of the Treasury Department’s and Federal Reserve’s Pandemic Response.” Testimony before the Financial Services Committee of the US House of Representatives, March 23.

Powell, Jerome. 2021b. “Opening Remarks: Monetary Policy in the Time of COVID.” In *Economic Policy Symposium Proceedings: Macroeconomic Policy in an Uneven Economy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Ramey, Valerie A., and Sarah Zubairy. 2018. “Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data.” *Journal of Political Economy* 126, no. 2: 850–901.

Reis, Ricardo. 2021. “Losing the Inflation Anchor.” *Brookings Papers on Economic Activity*, Fall, 307–61.

Seliski, John, Aaron Betz, Yiqun Gloria Chen, U. Devrim Demirel, Junghoon Lee, and Jaeger Nelson. 2020. “Key Methods That CBO Used to Estimate the Effects of Pandemic-Related Legislation on Output.” Working Paper 2020-07.

Washington: Congressional Budget Office. https://www.cbo.gov/system/files/2020-10/56612-Key-Methods.pdf.

Spilimbergo, Antonio, Prachi Mishra, Gita Gopinath, Daniel Leigh, and Laurence Ball. 2021. "US Inflation: Set for Take-Off?" *VoxEU*. https://cepr.org/voxeu/columns/us-inflation-set-take.

Staiger, Douglas, James Stock, and Mark Watson. 1997. "How Precise Are Estimates of the Natural Rate of Unemployment?" In *Reducing Inflation: Motivation and Strategy*, edited by Christina Romer and David Romer. Chicago: University of Chicago Press.

Summers, Lawrence H. 2021. "The Biden Stimulus Is Admirably Ambitious. But It Brings Some Big Risks, Too." *Washington Post*, February 4.

Summers, Lawrence H., and Fareed Zakaria. 2022. "On GPS: Is a Recession Inevitable?" Interview, CNN, July 23, https://www.cnn.com/videos/tv/2022/07/23/exp-gps-0724-summers-on-whether-recession-is-inevitable.cnn; transcript https://transcripts.cnn.com/show/fzgps/date/2022-07-24/segment/01.

Verbrugge, Randal. 2021. "Is It Time to Reassess the Focal Role of Core PCE Inflation?" Working Paper 21–10. Federal Reserve Bank of Cleveland. https://www.clevelandfed.org/publications/working-paper/2021/wp-2110-time-to-reassess-focal-role-of-core-pce-inflation.

Yellen, Janet. 2019. "Former Fed Chair Janet Yellen on Why the Answer to the Inflation Puzzle Matters." Remark delivered at "What's (Not) Up with Inflation?" Washington: Hutchins Center on Fiscal and Monetary Policy, Brookings Institution, October 3.

Comments and Discussion

COMMENT BY

JASON FURMAN[1] "Understanding US Inflation during the COVID-19 Era" by Laurence Ball, Daniel Leigh, and Prachi Mishra is among the scariest macroeconomic papers written in 2022. It diagnoses much of the increase in inflation in the United States as reflecting labor market tightness—and its model highlights the potential challenge of wringing this inflation out of the system. While I have some quibbles with particular parts of the analysis, overall I find it a reasonable quantification of the situation facing the US economy as a result of the enormous shock and extraordinary relief provided during the COVID-19 period.

This comment makes six points.

1. I HOPE THE PAPER IS WRONG Ball, Leigh, and Mishra's paper is a "choose your own adventure" that does not take a strong stance on the key parameters. Instead, the authors provide a forecast that is conditional on a trajectory for the unemployment rate and assumptions about a variety of the parameters.

The paper usefully focuses on two critical parameters. The first is shifts in the Beveridge curve. In the pandemic period the Beveridge curve has shifted out dramatically as shown in figure 14. As a result, even though the unemployment rate in mid-to-late 2022 was about 3.5 percent, as it was before the pandemic, the job openings rate was around 2 percentage points

1. I am indebted to Wilson Powell III for his usual outstanding research assistance on this comment.

higher.[2] As a result, the labor market was much tighter than it was prior to the pandemic, with about 1.7 job openings for every unemployed worker, up from a ratio of 1.2 prior to the pandemic.

The question is whether the Beveridge curve will shift back on its own, with reduced labor demand resulting in lower openings without rising unemployment. This is not something that has happened before (Blanchard, Domash, and Summers 2022), but then again neither have we seen such an abrupt and large shift out in the Beveridge curve.

The authors speculate about possible sources of this dramatic shift but neither they nor anyone else has a convincing story for why it has shifted so much. One possibility is a temporary response to the dramatic adjustments of the labor market during the pandemic period, for example, people finding new employers to satisfy their changed job preferences, such as for working from home. Under this possibility, once the labor market returns to normal the Beveridge curve would shift all the back to where it was before the pandemic.

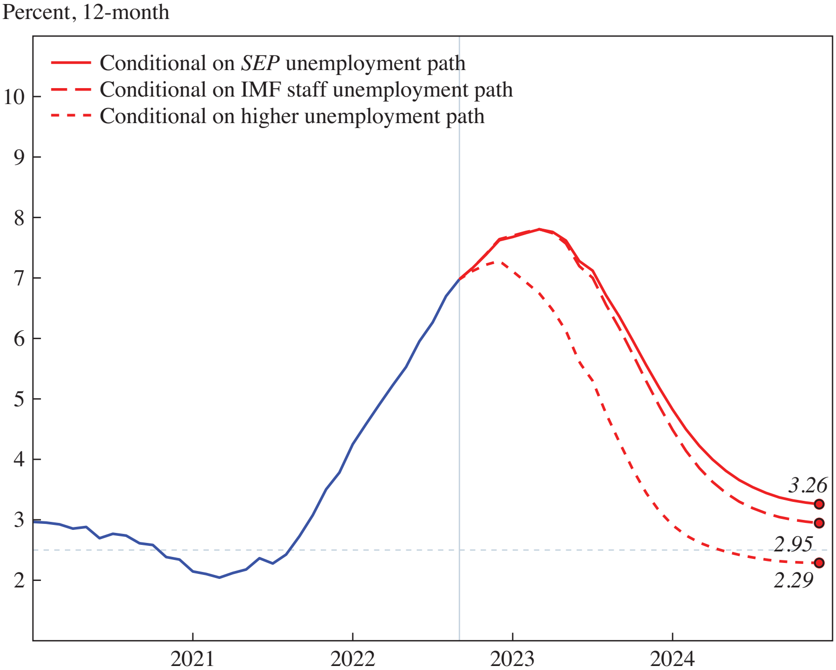
The fact that the Beveridge curve has shifted only a fraction of the way back in the year since COVID-19 became more normalized, expanded unemployment insurance expired, and schools reopened, however, suggests that it would not be reasonable to assume the Beveridge curve shifts all of the way back. In my scenarios I will assume, arbitrarily, that the Beveridge curve shifts two-thirds of the way back to where it was prior to the pandemic. This means some "immaculate" reduction in openings without a rise in the unemployment rate is possible, but that it would not return the economy all the way to where it was prior to the pandemic.

The second key issue is how inflation expectations evolve. The authors assume that long-run expectations update based on actual inflation. I assume that they are as anchored as they were prior to the pandemic ($\gamma = 0.991$ in the authors' model) but also that they exogenously shift halfway back to where they were pre-pandemic independent of the effect of actual inflation.

Finally, I follow the authors in assuming that going forward the headline shock is zero. In my comment at the conference in September, I assumed a cumulative −1 percentage point headline shock with headline Consumer Price Index (CPI) cumulatively lower than median inflation over the five remaining months of 2022. In the two months since the conference, this

2. Bureau of Labor Statistics, "Labor Force Statistics from the Current Population Survey (LNS14000000)," https://data.bls.gov/timeseries/LNS14000000; "Job Openings and Labor Turnover Survey (JTS000000000000000JOR)," https://data.bls.gov/timeseries/JTS000000000000000JOR.

Figure 1. Inflation Conditional on Different Unemployment Paths



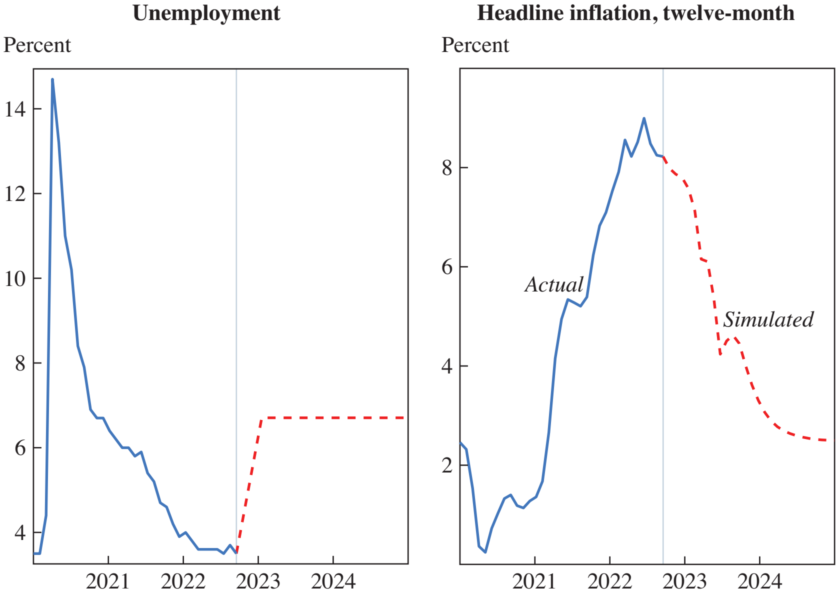
Source: Author's calculations.

entire shock has already happened. With this shock already incorporated into the updated data, I do not assume any further adjustment going forward.

The results of these assumptions are shown in figure 1. If unemployment follows the path in the September *Summary of Economic Projections* (*SEP*), maxing out at 4.4 percent in 2023, the median inflation rate would come down to 3.26 percent, equivalent to about 2.75 percent for the Federal Reserve's personal consumption expenditures (PCE) inflation target. If unemployment rises further to the 5.4 percent assumed by the International Monetary Fund (IMF) staff, inflation would only slightly exceed the Federal Reserve's target. Finally, if the unemployment rose to 7.5 percent for two straight years, as hypothesized by Lawrence Summers, inflation would fall slightly below the Federal Reserve's target by the end of 2024.

Overall to get inflation down the Federal Reserve's target under these assumptions would require the unemployment rate to be 6.7 percent for 2023 and 2024 as shown in figure 2.

Figure 2. Base Case for Unemployment Required to Hit the Federal Reserve's Inflation Target

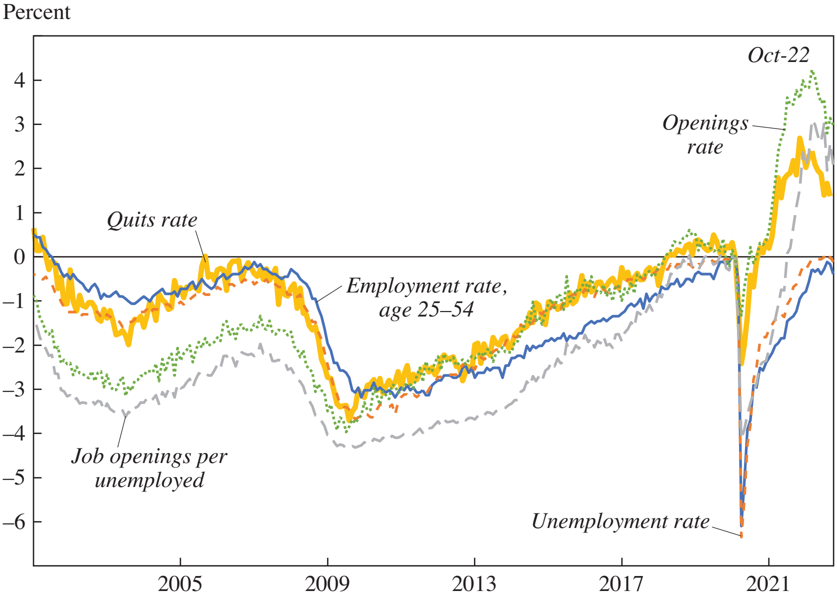


Source: Author's calculations.

2. *V/U* (OR *U/V*) IS AN IMPORTANT SLACK VARIABLE In recent years there has been an increased interest in a range of measures of labor market tightness that go beyond the unemployment rate. Two of the leading variables that have been advanced are quits and job openings (Furman and Powell 2021). There is a historical basis for this focus, but it came into strong relief over the last year because the unemployment rate was showing excess slack relative to the pre-pandemic labor market even while other measures like quits and openings showed a dramatically tighter labor market, as shown in figure 3, which normalizes a range of measures of labor market slack to zero prior to the pandemic with a standard deviation of one.

In theory, slack could be described as a function of the unemployment rate, the openings rate, and the quit rate, f (unemployment rate, openings rate, quit rate). Or it could be a function of the difference between the unemployment rate and the non-accelerating inflation rate of unemployment (NAIRU) with a time-varying NAIRU that depends on shifts in the Beveridge curve: f (unemployment rate – NAIRU [openings, quits]).

Figure 3. Measures of Labor Market Tightness



Sources: Bureau of Labor Statistics; Indeed Hiring Lab via Macrobond; author's calculations.

Note: Measures standardized using standard deviation from 2001 through 2018 and indexed to equal zero in February 2020. Prime-age employment is the share of the civilian population age 25–54 that is employed. Unemployment rate is the U-3 unemployment rate. The quits rate is quits divided by total nonfarm employment. The openings rate is openings divided by the sum of total nonfarm employment and openings. Job openings for October 2022 are estimated based on Indeed Hiring Lab job postings. The unemployment rate is plotted so that higher values correspond with a greater degree of labor market tightness, consistent with other measures.

The authors simplify all of this into a parsimonious single variable: slack = f (openings/unemployment) or $f(V/U)$.

Running a basic Phillips curve with different inflation concepts as the dependent variable $Inflation_{t \text{ to } t+4q} = \beta_0 + \beta_1 * Slack_t + \varepsilon_t$, the best predictor is actually the inverse of the authors' variable—the number of unemployed per job opening as shown in table 1. (Note that none of the slack variables are very good at explaining overall inflation which is very sensitive to exogenous shocks in energy and food prices.)

3. MEDIAN CPI IS THE RIGHT MEASURE OF INFLATION The authors argue, convincingly in my view, that median CPI is likely the right measure of inflation. In particular it has three desirable properties: (1) It is less volatile than core CPI. Over the last two years, for example, core CPI has been

Table 1. Adjusted R^2 in Phillips Curve Regressions for CPI

	Overall	*Ex food and energy*	*Trimmed mean*	*Median*
Unemployed per job opening	–0.01	0.42	0.30	0.68
Quits rate	0.01	0.41	0.35	0.67
Unemployment rate	–0.01	0.33	0.27	0.56
Job openings per unemployed	–0.01	0.29	0.19	0.46
Openings rate	–0.01	0.28	0.13	0.43
Prime-age employment rate	0.03	0.22	0.28	0.40

Sources: Bureau of Labor Statistics via Macrobond; author's calculations.

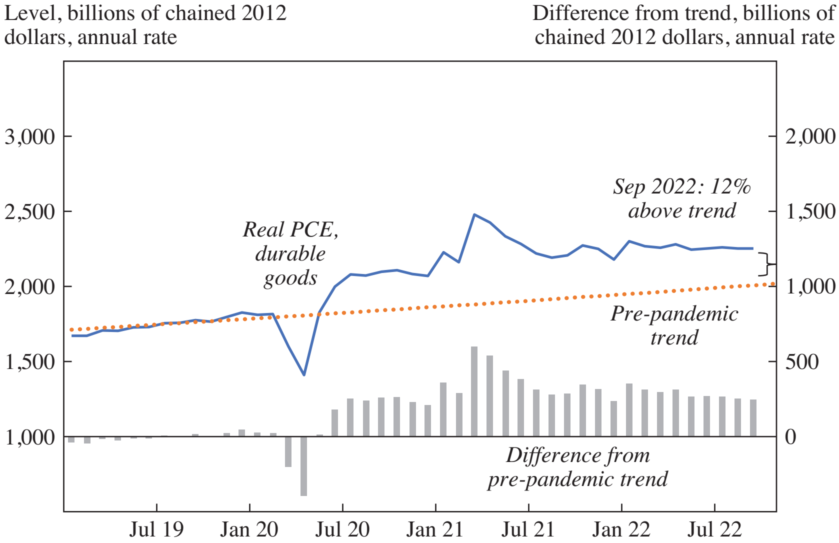
volatile as components like used cars have had outsized increases and decreases. By flexibly excluding outliers, median CPI is much less volatile. (2) It is a reasonable univariate predictor of future inflation. The median CPI provides as good, or perhaps a better, signal for future inflation as any other measure of underlying inflation. (3) Median CPI is much more predictable from labor market variables. Table 1 shows that *every* measure of labor market slack does a better job predicting median inflation than any other inflation concept. As such it appears to be effectively picking up "cyclically sensitive inflation" in the sense of Stock and Watson (2020).

The biggest criticism of the median CPI has been that shelter plays an outsized role in it.[3] The Federal Reserve Bank of Cleveland, which calculates the widely used median CPI, reduced the importance of shelter by dividing the component into four regions. Nevertheless, shelter in one of the regions is still the median category about half of the time. As the authors argue, it is not clear why this should bother us. Shelter is only the median item because half of the items (on a weighted basis) are above and below it. Moreover, the median of anything excluding the median is generally very close to the median assuming a smooth distribution. The fact that median CPI works so well empirically suggests these concerns are largely unfounded, although given the lags in the translation of spot rents to all rents, there is good reason to also keep an eye on other measures of underlying inflation.

4. HEADLINE SHOCKS REFLECT AN UNKNOWN COMBINATION OF SUPPLY AND DEMAND The authors develop a concept called "headline shocks" that is

3. Note that shelter is about one-third of the CPI but housing is only about one-sixth of the PCE price index. So this is a smaller issue for the median PCE. The paper, however, is focused on the median CPI. See Bureau of Labor Statistics, "Consumer Price Index," table 1 (2019–2020 Weights), "Relative Importance of Components in the Consumer Price Indexes: U.S. City Average, December 2021," https://www.bls.gov/cpi/tables/relative-importance/2021.htm.

Figure 4. Real Personal Consumption Expenditures, Durables



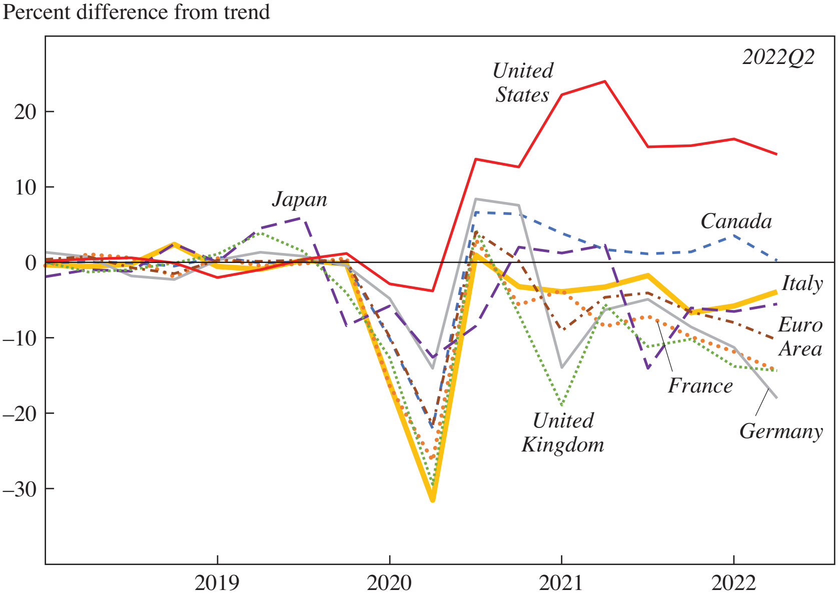
Sources: Bureau of Economic Analysis via Macrobond; author's calculations.
Note: Pre-pandemic trend based on log-linear regression for January 2018 to December 2019.

the difference between headline CPI and median CPI. Although they are agnostic about the interpretation of this headline shock, in general they lean into understanding it as a supply shock. This is problematic because unlike the difference between headline CPI and CPI excluding food and energy, this headline shock is really more about changes in the skewness of the CPI that are difficult to interpret.

Most of the measures the authors use to assess supply could just as easily be interpreted as demand. The Federal Reserve Bank of New York's Global Supply Chain Pressure Index, for example, records the difference between supply and demand—which is why it showed a rapid improvement in supply chains in the second half of 2008 when demand collapsed.

Consumption patterns skewed toward goods also appear to reflect demand as much if not more than supply. The big increase in consumer durables spending occurred when the economy and the service sector were rapidly reopening with the initially successful rollout of vaccines. As shown in figure 4, consumer durable spending was higher in June 2021 (when the economy was largely reopened) than it was in December 2020 (when the economy was much more closed). Moreover as shown in figure 5, goods spending soared in the United States in the face of massive fiscal stimulus

Figure 5. Real Durable Goods Expenditure Relative to Trend



Sources: Organisation for Economic Cooperation and Development via Macrobond; author's calculations.
Note: Pre-pandemic trend based on log-linear regression for 2018:Q1 to 2019:Q4. Euro area excludes Cyprus.

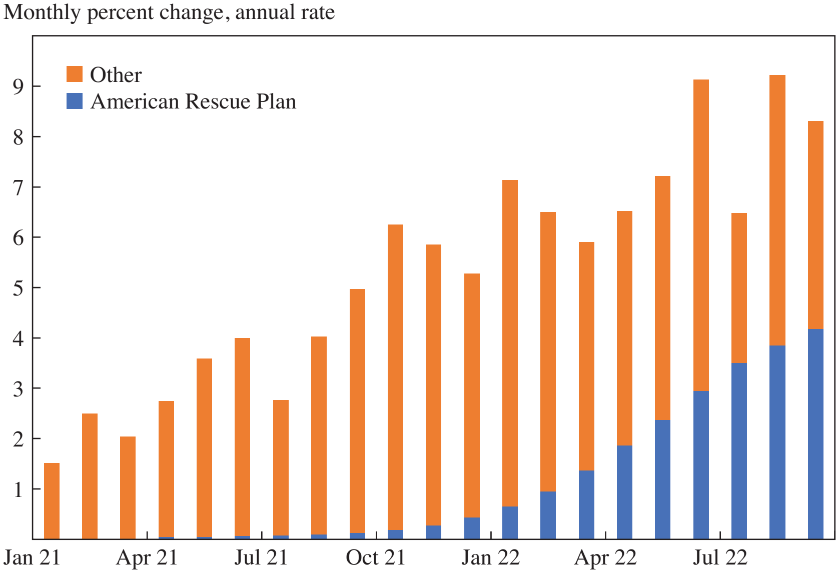
while not increasing above trend in other economies that were generally slower to reopen their service sectors.

5. THE PAPER MAY NEGLECT NONLINEARITIES AND TIMING EFFECTS FROM THE AMERICAN RESCUE PLAN The paper finds a relatively small effect of the American Rescue Plan on inflation in 2021 but a growing effect over the course of 2022 as shown in figure 6. This timing is the result of the lag structure assumed in their model: it takes some time for the American Rescue Plan to raise *V/U*, and then much of the effect that higher *V/U* has on inflation occurs over the following year. As a general matter this may be a reasonable lag structure, but for a massive change like the American Rescue Plan it is considerably less plausible that the effects were so small in 2021.

An alternative perspective on inflation, instead of modeling how stimulus affects the labor market and then how the labor market affects inflation, is to just go straight from the effect of stimulus on nominal GDP and then divide that impact into a price effect and an output effect (Furman 2022).

Specifically, using standard multipliers, output by 2021:Q3 would have been expected to be 4.8 percent above pre-pandemic estimates of potential.

Figure 6. Median CPI Inflation



Source: Author's calculations.

Moreover, pre-pandemic estimates of potential were unlikely to be a reasonable estimate of potential in 2021 because of reduced immigration, the time it takes to get people back into the labor force, forgone research and investment, and the lingering effects of other disruptions. To avoid inflation the economy would have needed to operate dramatically above potential in 2021. More realistically, the economy operated roughly at its potential with all of the additional nominal GDP showing up in the form of higher prices. This effect is much more immediate, occurring when the additional spending happened in 2021, not delayed to 2022 and operating through the labor market.

6. HOPE IS NOT A STRATEGY: IMPLICATIONS FOR MONETARY POLICY Finally, I would take three policy implications from the authors' model:

First, de-anchoring inflation expectations is costly. To the degree that acting more aggressively earlier keeps inflation expectations in check, that could lower the total cumulative jobs cost of achieving any given inflation goal. Specifically, table 2 shows what amount of unemployment would be needed in 2023 and 2024 (or the point years of added unemployment) to get the Federal Reserve's PCE inflation target down to 2 percent under various scenarios for expectations. To the degree that expectations are

Table 2. Inflation Expectations

	Unemployment in 2023 and 2024 needed for 2 percent PCE inflation	*Point years of added unemployment*
$\gamma = 0.90$	9.0	11.3
$\gamma = 0.945$ (1985 – 1998)	8.5	10.1
$\gamma = 0.991$ (2009 – 2019)	7.7	8.5
$\gamma = 0.991$ + 0.3 pp exogenous reduction	6.7	6.4
Revert to 2.2	4.9	2.5

Source: Author's calculations.

Table 3. Unemployment Increases Required for Different Inflation Targets

PCE inflation at end of 2024	*Unemployment in 2023 and 2024*	*Point years of added unemployment*	*Sacrifice ratio*
2.0	6.7	6.4	
			8.5
2.5	4.7	2.1	
			2.5
3.0	4.1	0.9	
			1.0
3.5	3.9	0.4	
			0.6
4.0	3.7	0.1	

Source: Author's calculations.

less anchored (as they were in the past) or do not exogenously decline above and beyond learning from actual lower inflation, the result is a much higher unemployment rate needed to control inflation—possibly as high as 9 percent.

Second, there is likely no way to get inflation down without at least a period of higher unemployment, likely above 4.5 percent—which would correspond to a recession. It takes special edge-case assumptions, like a full return of the Beveridge curve to its pre-pandemic relationship, for this to happen.

Third, the cost of lowering inflation is nonlinear and is much higher to lower inflation from 3 percent to 2 percent than it is to lower inflation from 4 percent to 3 percent as shown in table 3. This might complement other, longer-term reasons why a higher inflation target might be desirable. Of course, it is very tricky for the Federal Reserve to try to keep inflation expectations anchored if there is any reality to or perception of its shifting to a higher inflation target. Achieving a higher inflation target might be politically and practically impossible, but this analysis increases the desirability of achieving it.

Overall the paper makes an important contribution to both our understanding of the sources of inflation in the pandemic period as well as helping to guide us out of it—while giving us some key metrics to look at to understand inflation and its sources in the future.

REFERENCES FOR THE FURMAN COMMENT

Blanchard, Olivier, Alex Domash, and Lawrence H. Summers. 2022. "Bad News for the Fed from the Beveridge Space." Policy brief. Washington: Peterson Institute for International Economics. https://www.piie.com/publications/policy-briefs/bad-news-fed-beveridge-space.

Furman, Jason. 2022. "Why Did (Almost) No One See the Inflation Coming?" *Intereconomics* 57, no. 2: 79–86.

Furman, Jason, and Wilson Powell III. 2021. "What Is the Best Measure of Labor Market Tightness?" Blog post, Realtime Economics, Peterson Institute for International Economics, November 22. https://www.piie.com/blogs/realtime-economic-issues-watch/what-best-measure-labor-market-tightness.

Stock, James H., and Mark W. Watson. 2020. "Slack and Cyclically Sensitive Inflation." *Journal of Money, Credit and Banking* 52, no. S2: 393–428.

COMMENT BY

AYŞEGÜL ŞAHIN The COVID-19 pandemic which started in early 2020 resulted in a deep but brief recession. The unemployment rate rose from 3.5 percent to 14.7 percent from February 2020 to April 2020.[1] After the drastic drop in economic activity, the economy rebounded, and inflation, which had been dormant for two decades, flared up briskly with the core Consumer Price Index (CPI) inflation rising from 1.4 percent in January 2021 to 6 percent in January 2022.[2] This rapid rise in inflation initially was attributed to mostly transitory factors such as the shift in the composition of consumption from services to goods and supply chain disruptions reflecting pandemic-related factors. However, inflation turned out to be more persistent than initially assumed with the core CPI inflation still printing at 6.6 percent as of September 2022.

In this timely piece, Ball, Leigh, and Mishra examine the drivers of this recent surge in inflation and present projections for the medium-term

1. Bureau of Labor Statistics, "Labor Force Statistics from the Current Population Survey (LNS14000000)," https://data.bls.gov/timeseries/LNS14000000.

2. Bureau of Labor Statistics, "CPI for All Urban Consumers: All Items Less Food and Energy in U.S. City Average, 12-Month Percent Change (CUUR0000SA0L1E)," https://data.bls.gov/timeseries/CUUR0000SA0L1Es.

inflation outlook. They find that the rapid tightening of the labor market and the pass-through of past shocks to headline inflation to core inflation account for the run-up in inflation. They relate the headline shocks—defined as the difference between the total and core inflation—to increases in energy prices and backlogs of orders for goods and services. Lastly, the authors simulate the future path of inflation for alternative paths of the unemployment rate and argue that inflation is likely to remain above the Federal Reserve's inflation target unless unemployment rises by more than the Federal Reserve projects.

The reemergence of inflation after two decades of muted price increases is one of the key macroeconomic problems that we are facing as we approach the end of 2022. The Federal Reserve has been on a rapid tightening cycle, not seen since the early 1980s, to curb inflation and bring it back to its mandate-consistent level. With inflation remaining persistently high despite the 3 percentage points rise in the federal funds effective rate between March and November 2022, inflation will be our main focus of attention for years to come. Against this backdrop, the authors provide a detailed account of drivers of inflation and discuss the challenges we likely face going forward. This comment reviews and interprets the authors' findings and suggests new directions for research.

FRAMEWORK Ball, Leigh, and Mishra develop a multistep regression framework to decompose the surge in inflation and use their framework to provide projections under different assumptions. The multistep regression framework helps to introduce different potential drivers of high inflation despite not providing a clear decomposition between supply and demand factors.

It is useful to first review the regression framework to facilitate the interpretation of the findings and discuss their robustness. The paper starts with a simple, commonly used decomposition of observed inflation:

$$\pi_t = \pi_t^C + \pi_t^H, \tag{1}$$

where π_t is headline inflation, π_t^C is core inflation, and π_t^H is the deviation between headline and core inflation. It is important to note that while the authors refer to π_t^H as headline shocks, it represents the deviation between headline and core inflation and cannot be interpreted as an exogenous shock.

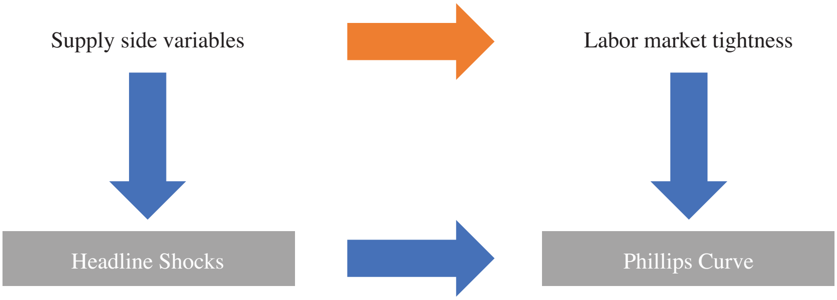
Step 1: Phillips curve regression. The first step is to run a Phillips curve–style regression which links core inflation to expected inflation, vacancy-to-unemployment ratio (V/U), and headline inflation shocks as well as quadratic and cubic terms. Specifically, the authors choose the following specification:

$$\pi_t^C - \pi_t^* = c + \underbrace{K_1 \frac{V_t}{U_t} + K_2 \left(\frac{V_t}{U_t}\right)^2 + K_3 \left(\frac{V_t}{U_t}\right)^3}_{\text{market tightness}} + \underbrace{\eta_1 \pi_t^H + \eta_2 \left(\pi_t^H\right)^2 + \eta_3 \left(\pi_t^H\right)^3}_{\text{headline shocks}} \tag{2}$$

where π_t^* is the expected inflation; $\frac{V_t}{U_t}$ is vacancy-to-unemployment ratio, often referred to as labor market tightness; and π_t^H is headline inflation shocks. For core inflation, the authors use the median CPI inflation rate published by the Federal Reserve Bank of Cleveland and use the Society of Professional Forecasters ten-year-ahead inflation expectations as a measure of π_t^*. Since the Job Openings and Labor Turnover Survey (JOLTS) starts only in 2000, the authors use the vacancy measures developed by Barnichon (2010), which combine the help wanted index with the JOLTS to construct a longer time series for vacancies. The authors capture the core inflation gap (median minus expected inflation) with four-quarter or twelve-month averages of *V/U* and headline shocks. The analysis focuses on the 1985–2022:Q3 period and does not use the data before 1985. This step picks up the co-movement between market tightness measures and inflation. In addition, the headline shocks, which are larger when the headline inflation deviates more from core inflation, could affect the core inflation gap.

Step 2: headline inflation regressions. The second step in analyzing inflation fluctuations is to interpret the deviation between headline and core inflation as headline inflation shocks and relate them to the recent developments in the macroeconomy. While the authors refer to the difference between headline and core inflation as headline inflation shocks, they do not specifically identify these shocks. Instead, in the second step of their regression framework they identify some variables that correlate with these deviations. They argue that shifts in either industry supply or industry demand could affect the headline inflation but do not attempt to decompose these into supply and demand channels. Instead, they find that changes in energy prices, backlogs of orders for goods and services, and changes in auto-related prices are positively related to headline inflation shocks. Clearly, these variables are endogenous to shifts in demand, shifts in composition of demand, and labor supply constraints.

Figure 1. Simple Diagram of the Multi-regression Framework



Source: Author's compilation.

Step 3: decomposing the surge in inflation. Equipped with two multivariate regressions, the authors then use the two reduced-form relationships consecutively to decompose the 6.9 percentage point rise in headline inflation, from 1.3 percent in December 2020 to 8.2 percent in September 2022. In particular, they first determine the contributions to the rise in inflation of higher expected inflation, higher levels of the vacancy-to-unemployment ratio, and headline shocks. Then they use the headline inflation regression to determine the shares of headline shocks attributed to energy price shocks, backlogs, and auto price shocks. They find that the direct and pass-through effects of headline inflation shocks account for about 4.6 percentage points of the 6.9 percentage point rise in twelve-month inflation. Most of this 4.6 percentage point total reflects energy price shocks and backlogs of work, with total contributions of 2.7 and 1.7 percentage points, respectively. The contribution of the vacancy-to-unemployment ratio to the rise in twelve-month inflation is 2 percentage points, nearly a third of the total inflation increase. A rise in expected inflation accounts for 0.5 percentage points. One caveat is that while market tightness does not seem to be the major factor in accounting for the rise in inflation, its importance seems to be rising over time. The results are shown in the authors' figure 12.

IS THE MULTISTEP MULTI-REGRESSION APPROACH REASONABLE? The appeal of the multi-regression framework is its simplicity. It helps connect each driver to core inflation either through the Phillips curve regression or through its direct or indirect effect through headline inflation shocks. Figure 1 shows how the multi-regression framework isolates the role of headline shocks by assuming that labor market tightness is not affected by those shocks. This assumption allows the authors to run repeated regressions and decompose the role of headline shocks and market tightness separately. However, it is likely that headline shocks had direct effects on the labor market, making it harder to interpret the decomposition. However, empirical evidence suggests that supply chain disruptions that the authors interpret as headline shocks had effects on the labor market. Amiti and others (forthcoming) document a big rise in imported input prices in the 2020:Q2–2022:Q1 period which coincided with a stark rise in wages. Import prices (excluding petroleum) increased by 6.7 percent during this period while the Employment Cost Index (ECI) increased by 4.1 percent. This is in contrast to the 2009:Q4–2019:Q3 period when the change in import prices was negligible and the ECI grew by 2.2 percent. They argue that, in normal times, firms can substitute between labor and imported intermediate inputs, thus cushioning any cost shock due to one of the two factors. This substitution mechanism has been highlighted by Feenstra and others (2018) and Elsby, Hobijn, and Şahin (2013). When labor costs go up, firms can outsource production to other countries and import intermediate inputs.

Over the past decades, US inflation has become more closely linked to global factors, as foreign competition and firms' ability to outsource have weakened the link between wage pressures and prices in the United States, as argued by Forbes (2019) and Obstfeld (2020). However, this substitution channel was less operational in the post-COVID-19 economy due to the large and simultaneous inflationary shocks to both labor and intermediate inputs. Moreover, US firms become less concerned about losing market share to foreign competitors when the shock is global in nature, raising their pass-through of cost shocks into prices. Amiti and others (forthcoming) use weekly earnings from the Quarterly Census of Employment and Wages (QCEW) for six-digit North American Industry Classification System (NAICS) industries in 2013–2021 and the Producer Price Index (PPI) for 527 six-digit NAICS industries and show that rising input prices are associated with increasing wages across industries, especially in 2021. They find that about one-third of the uptick in wage inflation can be attributed to the supply chain problems. In addition, they show in more detail that rising import prices triggered a shift away from imported intermediaries to domestic labor and wages and employment. This substitution channel suggests that headline shocks that the authors identify likely have affected the labor market and their true contribution on inflation is likely to be larger through their effects on labor market tightness. This orthogonality assumption is also important for the projections in the paper. Since market tightness is affected by headline shocks, when their effects subside, there should be a direct negative effect on market tightness as well.

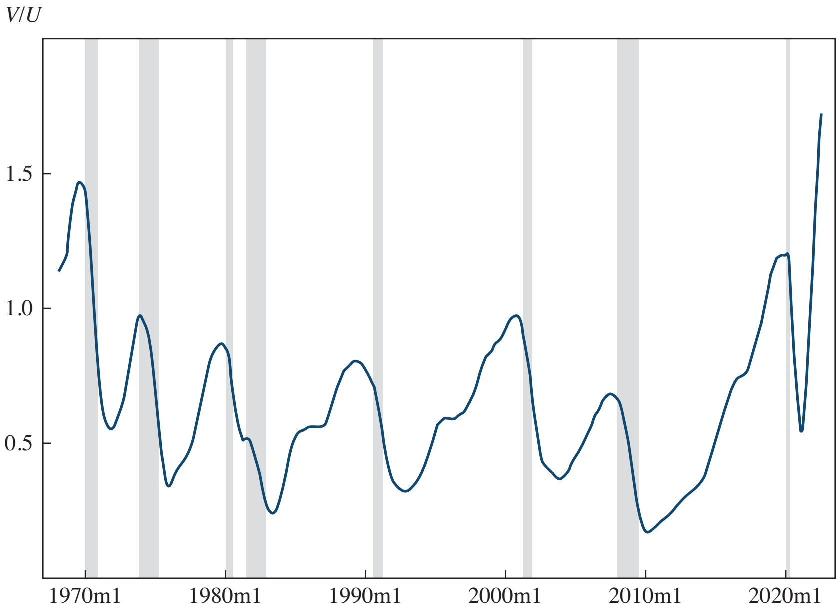
INTUITION FOR HIGHER-ORDER TERMS Another concern I have is the use of quadratic and cubic terms in the Phillips curve regressions which seem to capture the recent inflation behavior well. Paradoxically, there is little evidence of nonlinearity for wage inflation, especially when the authors include the pandemic period in their sample. I find these results hard to interpret. These higher-order terms seem significant and quantitatively important for price inflation, but the authors do not provide an economic explanation for why they would matter so much. The mechanism behind this nonlinearity remains unexplored but is vital for inflation projections. For example, the authors' figure 10 shows the wage inflation and price inflation gaps as a function of vacancy-to-unemployment ratio. In contrast to the results for price inflation, the estimated effect of market tightness on wage inflation is approximately linear. This disconnect between wage and price inflation makes the importance of higher-order terms of *V/U* on price inflation more puzzling since they do not seem to originate from wage pressures.

Since the paper's preferred measure of labor market slack is the vacancy-to-unemployment ratio, connection to the vast search and matching theory can help provide some intuition. For example, a recent literature argues that job-to-job transitions capture wage pressures better than the unemployment-to-employment transition rate by analyzing the predictive power of the unemployment rate, the unemployment-to-employment transition rate, hires from nonparticipation, and job-to-job transitions (Faberman and Justiniano 2015; Faccini and Melosi 2021; Karahan and others 2017; Moscarini and Postel-Vinay 2017, 2022). These papers argue that behavior of wages is better captured by job-to-job transitions than the unemployment rate. Since job-to-job transitions constitute a higher fraction of hires during tight labor markets, they might create the type of nonlinearities the authors identify. The underlying reasons for the nonlinearity remain an open and important issue for future research.

IS VACANCY-TO-UNEMPLOYMENT A PANACEA FOR THE PHILLIPS CURVE? That economists have long been pursuing the perfect measure of slack and emphasis on labor market tightness is nothing new. For example, George Perry, in one of the first Brookings papers, wrote: "For instance, many (including myself) argue that what matters is the difference between available jobs and available employees to fill those jobs" (1970, 412).

I like that the authors use the vacancy-to-unemployment ratio as the measure of labor market tightness, but using tightness alone does not solve the trend and compositional issues that other measures of slack are criticized for. This is clear in the historical time series of the vacancy-to-unemployment ratio plotted in figure 2. The vacancy-to-unemployment ratio averaged

Figure 2. Historical Evolution of Vacancy-to-Unemployment Ratio



Sources: Author's calculations; Bureau of Labor Statistics JOLTS.

at 0.70 in 1970–1979. In this period the core CPI inflation increased by 5.1 percentage points. In the January 2021 to September 2022 period, the vacancy-to-unemployment ratio averaged 1.06 and the core CPI inflation increased by 5.2 percentage points. Interestingly, during the Great Recession, which was characterized by subdued inflation, the US labor market was tighter than in the 1970s according to the measure used in the paper. The authors also show that their Phillips curve regression does not fit the 1970s well. Even the use of higher-order terms of *V/U* in the Phillips curve regressions does not capture the evolution of price inflation in the 1970s.

One potential problem about the vacancy series is that historical data and post-2000 data come from different data sources. The historical help wanted series and the more recent JOLTS data are very different, which makes it harder to interpret the level of *V/U* over time. But this disconnect in the historical data applies to observations between 1985 and 2000 as well. The second issue is the change in trend unemployment over time. Unemployment has trended down since the 1980s and measures of the natural rate of unemployment or NAIRU take this trend change into account to estimate cyclical changes in the unemployment rate. Using *V/U* alone without considering the changes in trend unemployment naturally inherits the same issues one encounters when using the unemployment rate as a measure of slack. Lastly, Abraham, Haltiwanger, and Rendell (2020) developed a generalized measure of labor market tightness which takes into account intensive and extensive margins of search activity on both demand and supply sides of the labor market and show that their measure captures the hiring process in the US economy better than the standard measure of labor market tightness. Their measure could potentially help explain why the fit is so bad for the 1970s.

Unfortunately, the authors do not investigate the economic mechanisms and measurement issues that might account for this poor fit, ignoring the 1970s in their analysis and only using post-1985 data. This choice, of course, comes at the expense of ignoring the only other high-inflationary episode in the last fifty years.

THE BEVERIDGE CURVE AND THE ROLE OF JOB LOSS IN A SOFT VS. HARD LANDING

Inflation projections in the paper are based on a log-linear relationship between tightness and unemployment in the form of

$$\frac{V}{U} = aU^{b-1}. \tag{3}$$

The authors estimate this functional form with pre-pandemic data on unemployment and vacancies which they refer to as the pre-pandemic Beveridge curve. Then they focus only on the April 2020–August 2022 data and estimate a post-pandemic Beveridge curve. They rely on these estimates to convert the unemployment projections to V/U with and without shifts in the Beveridge curve. The crucial assumption for this approach is that there is a one-to-one mapping between the unemployment rate and tightness. This assumption ignores the accounting identity that captures the evolution of unemployment which implies:

$$U_{t+1} = s_t\left(1 - U_t\right) - f_t U_t, \tag{4}$$

where s_t is the inflow rate to unemployment and f_t is the outflow rate from unemployment.

Search and matching frictions are typically summarized by the matching function of the Cobb–Douglas form which links hires to unemployment U and vacancies V:

$$f = \frac{Hires}{U} = \frac{M\left(V, U\right)}{U} = A\left(\frac{V}{U}\right)^{\sigma}, \tag{5}$$

where A is the matching efficiency and σ is the elasticity of the matching function. The search and matching literature typically estimates the matching function using data on hires or job-finding rate, vacancies, and unemployment. Instead, the authors choose a functional form which makes it harder to compare their estimates with those in the literature. The flow steady state implies a Beveridge curve of the form:

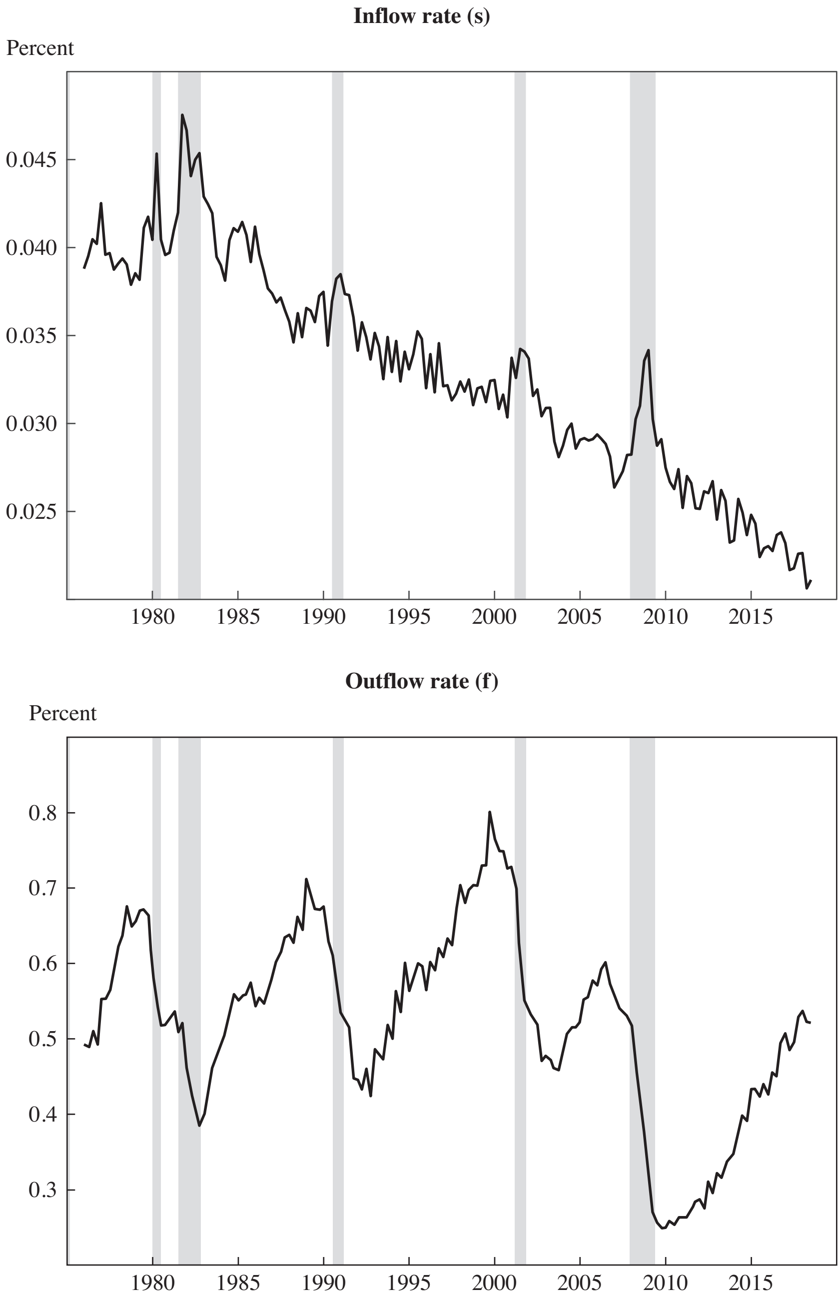
$$u = \frac{s}{s+f} = \frac{s}{s + A\left(\frac{V}{U}\right)^{\sigma}}, \tag{6}$$

as shown by Pissarides (1985). The implication of equation (6) is that the position of the Beveridge curve depends on the unemployment inflow rate, s. Increases in the inflow rate, which is associated with increases in layoffs and job destruction, shift the Beveridge curve out, implying a higher unemployment rate for the same level of vacancies. On the contrary, soft landings are associated with small increases in unemployment inflows. Figure 3 shows that the inflow rate increased sharply at the onset of deep recessions while it exhibited a muted response during mild recessions, such as the 1991–1992 and 2001 recessions which are interpreted as soft landings. On the contrary, the behavior of the outflow rate is very similar regardless of the severity of recessions. More importantly, contractionary monetary policy shocks tend to affect the unemployment inflow rate first. While the soft versus hard landing discussions in the paper focus on only one determinant of the Beveridge curve, V/U, the inflow rate is likely to be important in the near future.

CONCLUDING REMARKS Ball, Leigh, and Mishra provide a detailed account of inflation developments in the post-pandemic economy. They consider many different drivers of inflation and identify various interesting patterns. While this is a useful exercise to identify important channels, relying on a multi-regression framework likely would make the results less relevant as the economy goes through a new boom-bust cycle in the future. I am especially concerned about using different labor market indicators to explain different inflationary or disinflationary episodes as in the case of short-term unemployment to account for the inflation dynamics after the Great Recession. My preferred measure of labor market has been the unemployment rate.

In my view, a useful construct to gauge the unemployment-inflation trade-off is the so-called natural rate of unemployment, which is defined

Figure 3. Historical Evolution of Unemployment Inflow and Outflow Rates



Source: Author's calculations using data from Crump and others (2022).

as the unemployment rate such that, controlling for supply shocks, inflation remains stable. The natural rate of unemployment is affected by both business cycle fluctuations and secular factors. Furthermore, the unemployment-inflation trade-off is linked by the classical determinants of inflation, such as inflation expectations. To accommodate all of these facets, a comprehensive framework is required that uses a New Keynesian Phillips curve as well as detailed information on unemployment flows such as in Crump and others (2022). In this model, the natural rate is informed by wage and price inflation, inflation expectations, and changing secular factors. This micro-macro Phillips curve framework not only creates a clear link between the labor market and inflation, it also directly incorporates the movements in survey-based inflation expectations. A New Keynesian Phillips curve estimated with rich labor market data captures the joint behavior of unemployment, wage and price inflation, and inflation expectations in the 1960–2022 period very well with a time-invariant slope—estimated to be quite flat. Even if the slope of the Phillips curve is small in a forward-looking model, this does not necessarily imply a weak link between the unemployment gap and inflation. According to the micro-macro Phillips curve in Crump and others (2022), the natural rate of unemployment was around 5 percent before the onset of the pandemic and increased to 7 percent by mid-2022. This pronounced rise was primarily informed by strong wage growth rather than changes in inflation expectations. The model-based forecasts in Crump and others (2022) suggest that strong wage growth is likely to moderate only sluggishly, continuing to put upward pressure on inflation in the medium run. The model forecasts the unemployment rate to rise to around 5 percent by mid-2024 and the unemployment gap to narrow, bringing underlying inflation to around 2.8 percentage points—about 0.5 percentage points above its long-run trend. While episode-specific analysis could give helpful hints about recent developments in inflation, a model-based approach is likely to provide more enduring insights into determinants of inflation.

REFERENCES FOR THE ŞAHIN COMMENT

Abraham, Katharine, John Haltiwanger, and Lea Rendell. 2020. “How Tight Is the US Labor Market?” *Brookings Papers on Economic Activity*, Spring, 97–165.

Amiti, Mary, Sebastian Heise, Fatih Karahan, and Ayşegül Şahin. Forthcoming. “Inflation Strikes Back: The Role of Import Competition and the Labor Market.” *NBER Macroeconomics Annual*.

Barnichon, Regis. 2010. “Building a Composite Help-Wanted Index.” *Economics Letters* 109, no. 3: 175–78.

Crump, Richard K., Stefano Eusepi, Marc Giannoni, and Ayşegül Şahin. 2022. "The Unemployment-Inflation Trade-off Revisited: The Phillips Curve in COVID Times." Working Paper. https://www.dropbox.com/s/24osn6dp5dnrfgw/CEGS_Update_20221001.pdf.

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin. 2013. "The Decline of the U.S. Labor Share." *Brookings Papers on Economic Activity*, Fall, 1–52.

Faberman, Jason, and Alejandro Justiniano. 2015. "Job Switching and Wage Growth." *Chicago Fed Letter*, no. 337. Federal Reserve Bank of Chicago.

Faccini, Renato, and Leonardo Melosi. 2021. "Bad Jobs and Low Inflation." Working Paper 2020-09. Chicago: Federal Reserve Bank of Chicago. https://doi.org/10.21033/wp-2020-09.

Feenstra, Robert C., Philip Luck, Maurice Obstfeld, and Katheryn N. Russ. 2018. "In Search of the Armington Elasticity." *Review of Economics and Statistics* 100, no. 1: 135–50.

Forbes, Kristin J. 2019. "Inflation Dynamics: Dead, Dormant, or Determined Abroad?" *Brookings Papers on Economic Activity*, Fall, 257–319.

Karahan, Fatih, Ryan Michaels, Benjamin Pugsley, Ayşegül Şahin, and Rachel Schuh. 2017. "Do Job-to-Job Transitions Drive Wage Fluctuations over the Business Cycle?" *American Economic Review* 107, no. 5: 353–57.

Moscarini, Giuseppe, and Fabien Postel-Vinay. 2017. "The Relative Power of Employment-to-Employment Reallocation and Unemployment Exits in Predicting Wage Growth." *American Economic Review* 107, no. 5: 364–68.

Moscarini, Giuseppe, and Fabien Postel-Vinay. 2022. "The Job Ladder: Inflation vs. Reallocation." Working Paper. https://cpb-us-w2.wpmucdn.com/campuspress.yale.edu/dist/1/1241/files/2022/09/nominal-rigidities_SA.pdf.

Obstfeld, Maurice. 2020. "Global Dimensions of U.S. Monetary Policy." *International Journal of Central Banking* 16, no. 1: 73–132.

Perry, George L. 1970. "Changing Labor Markets and Inflation." *Brookings Papers on Economic Activity*, no. 3: 411–48.

Pissarides, Christopher A. 1985. "Short-Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages." *American Economic Review* 75, no. 4: 676–90.

GENERAL DISCUSSION James Stock commented that the ratio of the number of vacancies to the number of unemployed (*V/U*) and the natural rate of unemployment (U^*) are mathematically equivalent, as both quantities have just one time-varying slack parameter.

Robert Hall said that while *V/U* is generally a good measure of labor market tightness, during the COVID-19 pandemic a complication arose of laid-off workers subject to recall, who are not measured in *V* and so must be removed from *U*. He added that making such a correction would be feasible and yield more sensible results.

Austan Goolsbee argued that comparing quantities involving vacancies and unemployment (such as V/U) across many decades is problematic, due to changing definitions of who is considered unemployed versus not in the labor force, such as the consideration of disability.

Alan Blinder asked whether online job boards have increased the ease of posting a vacancy, leading to duplicate vacancies and thus an aggregate measure of vacancy that is inconsistent with past measures.

Ricardo Reis suggested that the paper modify its metric of inflation expectations, from the Society of Professional Forecasters' ten-year median inflation expectations to the University of Michigan Survey of Consumers' one-year mean inflation expectations. Reis argued that professionals as a population and medians as a statistic, which the paper uses, are too stable recently and contain very little signal. A survey of households (such as Michigan's) and a mean measure that is more affected by the answers of the tail would be preferable. Further, with reference to past work by Hazell and others, Reis added that the theoretical case for using long-run inflation expectations of ten years, as used in the paper, relies on including in the regression long-run unemployment expectations, which the paper does not do. Instead, one-year inflation expectations are consistent with the short-run unemployment measure used in the paper.[1]

Blinder criticized the definition of core inflation used by the authors, which excludes unpredictable components of inflation. He contended instead that core inflation was intended to measure components which may be affected by aggregate demand policies, primarily monetary policy. He observed that the paper incorrectly removed automobile prices from core inflation, even though monetary policy does directly affect the automobile market through interest rates on auto loans.

Robert Gordon asked why the authors had chosen a cubic functional form for their regression, which would imply that a low V/U leads to rapid disinflation, contrary to evidence from 2009 and 2010. He added that it was the failure of inflation to slow down in the presence of high unemployment after the Great Recession that had discredited the Phillips curve.

Justin Wolfers expressed skepticism of the paper's results, given the small number of data points corresponding to the large number of degrees of freedom available for curve-fitting. Wolfers listed a number of such degrees of freedom available to researchers, including the type of inflation

1. Jonathan Hazell, Juan Herreño, Emi Nakamura, and Jón Steinsson, "The Slope of the Phillips Curve: Evidence from U.S. States," *Quarterly Journal of Economics* 137, no. 3 (2022): 1299–344, https://doi.org/10.1093/qje/qjac010.

metric, the type and time horizon of inflation expectation, the choice of survey data source, the measure of labor market slack, the time variance of model coefficients, the model lag structure and nonlinearities, and the inclusion of supply shocks and regime shifts. He argued that given such a large number of degrees of freedom, any Phillips curve could be plausibly claimed to account for the observed path of inflation.

James Hamilton noted that the model used in the paper was highly nonlinear based on a cubic function of the ratio *V/U*. He suggested instead starting with a specification that is linear in the logs of the primitive variables *V* and *U*, similar to equations presented in Ayşegül Şahin's discussion paper, and then seeing if it was helpful to generalize this to a function that is quadratic in logs.

Hall contended that the paper's use of a complicated autoregressive specification was not necessary or explanatory and that it would be preferable to return to a more fundamental microeconomic view of inflation being a result of buyers and sellers agreeing to higher prices. Hall added that the Beveridge curve is not a structural object, in agreement with Şahin.

Emi Nakamura echoed the challenges of modeling specifications for a Phillips curve, particularly identification in a time series context.

Maurice Obstfeld argued that the present inflation scenario was not in the standard Phillips curve region and should not be modeled as such, in agreement with Jason Furman, and with reference to John Maynard Keynes's argument in his 1940 book titled *How to Pay for the War*.[2] Rather, Obstfeld said that present inflation in the United States should be understood to be a result of nominal demand exceeding nominal supply, due to a highly supply-constrained economy resembling postwar Europe or present-day Ukraine.

Nakamura asked the authors about the role the sectoral shocks and relative price shocks play in inflation, and especially how they affect future inflation projections. Nakamura suggested that when comparing the present inflation to that of the 1970s, a parallel of large relative price variability arises; this was a part of the reason for the belief that inflation would be transitory and that there would be a reduction in relative price shocks, connected to the issue of sectoral shifts and supply shocks.

Goolsbee stated that if inflation were caused by supply shocks, then the forecast of inflation is equivalent to a forecast of the likelihood of reversing the supply shocks. Further, Goolsbee argued that the fact that before the

2. John Maynard Keynes, *How to Pay for the War: A Radical Plan for the Chancellor of the Exchequer* (London: Macmillan, 1940), available at https://fraser.stlouisfed.org/title/6021.

COVID-19 pandemic, the unemployment rate was 3.5 percent with little inflation, but that later inflation increased when unemployment was at 6 percent, and suggested that inflation had a significant supply shock component even before looking for nonlinearity in the Phillips curve.

Claudia Sahm agreed with Goolsbee in arguing that the Phillips curve is inappropriate to the study of present inflation, as the Phillips curve is far better specified for demand shocks than to the supply shocks that underlie present-day inflation; using the Phillips curve may therefore lead us astray on how inflation and unemployment need to be addressed. She conceded that the paper made progress on understanding inflation and the Phillips curve in the context of supply shocks, with decompositions of the median CPI and headline inflation, and accounting for supply chain shocks.

Laurence Meyer stated that a nonlinear Phillips curve and a higher non-accelerating inflation rate of unemployment were persuasive conclusions of the empirical research.

Laurence Ball accepted Şahin's and Furman's critiques of the Phillips curve, particularly that historically attempts to model it have suffered from poor specification, although little could be done in response except to continue to improve modeling efforts. Ball added that either supply or demand shocks can cause relative shocks within sectors, and that the paper does not disaggregate the effects of supply from demand but instead focuses on how labor markets and relative sectoral shocks feed into inflation.

Goolsbee added that the fraction of inflation that is caused by energy will make a huge difference, although he agreed with Furman on the point of some measures of supply chain tightness possibly being measures of demand, not supply. Goolsbee suggested that the authors should have considered the effect of productivity on wages and prices, particularly given fluctuating productivity through the COVID-19 pandemic period.

Sahm argued that the paper did not adequately account for labor market shortages due to COVID-19 pandemic-imposed illness and mortality, which may take a long time to dissipate. As a comparison, she referenced an analysis of the effect of pandemics on the labor market.[3]

Meyer said that a regime shift, from a low-inflation to a high-inflation regime, should be a critical consideration in modeling inflation. An important component of this regime shift is the much higher importance of short-term inflation expectations in wage bargaining, relative to long-run inflation

3. Oscar Jorda, Sanjay R. Singh, and Alan M. Taylor, "Longer-Run Economic Consequences of Pandemics," working paper 2020-09, Federal Reserve Bank of San Francisco, https://doi.org/10.24148/wp2020-09.

expectations. Meyer added that adaptive inflation expectations should be considered.

Gordon emphasized that the Federal Reserve Board could choose an inflation target of 2 percent or 3 percent, and that sticking to a 2 percent target would be costly and warrants public discussion. Blinder responded that he did not expect the Federal Reserve to change the target.

Gordon stated that Furman's parameter choices were extremely optimistic; contrary to Furman's choices, inflation expectations are unlikely to be re-anchored to pre-pandemic levels and *V/U* is unlikely to return to two-thirds of the way back to pre-2019 levels.

Frederic Mishkin expressed pessimism regarding the near future of unemployment, arguing that the Federal Reserve may need to cause a serious recession to control inflation. Mishkin argued that the Federal Reserve has made two serious mistakes in addressing inflation—it abandoned its preemptive strategy, and it didn't specify a horizon for average inflation targeting. Due to these mistakes, inflation expectations are less anchored and Federal Reserve actions need to be tougher than they otherwise would be.

Reis disputed Gordon's view, instead stating that he was optimistic about inflation reducing going forward, noting that he observed a dramatic re-anchoring of expectations in the Michigan mean of one-year and five-year inflation expectation numbers in the preceding three-month period, maybe as a result of communications from the Federal Reserve becoming tougher on inflation.

Responding to points raised in the discussion papers, Ball noted that Şahin's outcomes were similar to the authors' outcomes. Ball agreed with Şahin's suggestion that *V/U* and headline shocks may be influenced by similar factors, such as the overheating which contributed to supply chain troubles. Ball added that he did not understand Şahin's argument that for the authors' two-stage regression to be valid, supply side variables, such as wages, must be uncorrelated with labor market tightness; however, he stated that the two quantities were not strongly correlated empirically.

JONATHAN A. PARKER
Massachusetts Institute of Technology

LAURA ERHARD
Bureau of Labor Statistics

JAKE SCHILD
Bureau of Labor Statistics

DAVID S. JOHNSON
National Academies of Sciences, Engineering, and Medicine

Economic Impact Payments and Household Spending during the Pandemic

ABSTRACT Households spent only a small fraction of their 2020 Economic Impact Payments (EIPs) within a month or two of arrival, consistent with pandemic constraints on spending, other pandemic programs and social insurance, and the broader disbursement of the EIPs compared to the economic losses during the early stages of the pandemic. While these EIPs did not fill an urgent economic need for most households, the first round of EIPs did provide timely pandemic insurance to some households that were more exposed to the economic losses from the pandemic. Households with lower liquid wealth entering the pandemic and those less able to earn while working from home raised consumption more following receipt of their EIP. While our measurement for later EIPs is not as reliable, our estimates suggest even less spending on average to the second and third rounds of EIPs. Our point estimates imply less short-term spending on average than in response to economic stimulus payments in 2001 or 2008. While our analysis lacks the power to measure longer-term spending effects, the lack of short-term spending contributed to strong household balance sheets as the direct economic effects of the pandemic on households waned.

Conflict of Interest Disclosure: Laura Erhard and Jake Schild are employees of the Bureau of Labor Statistics (BLS). BLS reviewed the research to ensure that the paper does not take a political stance and that it has correctly and accurately described and analyzed the consumer expenditure data that the BLS produces and releases to the public. 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, Fall 2022: 81–130

In response to the economic consequences of the pandemic, the United States government distributed three waves of Economic Impact Payments (EIPs) to American households. In March 2020, following the declaration of a national emergency, Congress passed the Coronavirus Aid, Relief, and Economic Security (CARES) Act. The act authorized more than $2 trillion of spending on programs that included the disbursement of roughly $300 billion in EIPs to the vast majority of Americans. In December 2020 with the pandemic continuing, the Coronavirus Response and Relief Supplemental Appropriations (CRRSA) Act authorized a second wave of roughly $150 billion in EIPs, and in March 2021, the American Rescue Plan (ARP) Act authorized a third round of just over $400 billion in EIPs.[1]

While these payment programs were modeled on stimulus payment programs that the government had implemented at the beginning of recessions in both 2001 and 2008, the economic situation in the pandemic was entirely different. The pandemic caused a large collapse in production as well as demand, as people—partly at the behest of the government—cut back on both producing and consuming goods and services that risked exposure to COVID-19. Thus, the EIPs were not intended to stimulate demand for consumption but rather to provide pandemic insurance, ensuring that people who had unexpectedly lost their livelihoods could continue to cover their consumption needs and financial obligations. The EIPs were not targeted to those who had lost their incomes, but were widely distributed, presumably for reasons of feasibility and expediency, as well as to get aid to people who were experiencing the impact of the pandemic but not eligible for aid through other programs.

In this paper, we study the responses of consumer spending to the arrival of the EIPs and evaluate the extent to which the EIPs provided widespread, urgently needed pandemic insurance. Focusing first on the spending response to the first round of EIPs, we estimate that the spending of the average household rose only a small amount over the couple of months following the arrival of their EIP, when compared to households that received later EIPs or did not receive EIPs at all, suggesting that the typical recipient was not in dire need of the EIP. We do, however, find larger spending responses both for those households with low levels of ex ante liquid wealth and for those more reliant on earnings from jobs that could not be done from home. While our data do not measure the arrival of the second and third rounds of EIPs as well as they do the first round, our estimates suggest even lower

1. The CRRSA Act was included as a part of the Consolidated Appropriations Act of 2021, which was signed into law on December 27, 2020.

average, short-term spending responses to these final two rounds. Finally, we find some evidence of spending over the three months following our initial short-term spending estimates but lack the statistical power to measure the spending effects of any round of EIPs over a longer period; we can only conclude that the lack of short-term spending contributed to strong household balance sheets as the economic effects of the pandemic waned following the three rounds of EIPs.

Our results are based on analysis of the Consumer Expenditure (CE) Interview Survey. We measure the average response of consumer spending to the receipt of an EIP using variation across households in receipt, in amount conditional on receipt, and in when they received a payment. As a baseline, we compare our estimates of spending to those reported in Johnson, Parker, and Souleles (2006) and Parker and others (2013) for the 2001 and 2008 tax payments using exactly the methodology employed in these papers. But there are substantial differences not only between program goals but also between the structure of these payment programs and the structure of the EIP programs. The EIPs were disbursed more widely, more rapidly (and so less drawn out over time), and more often by direct deposit, and rounds one and three were larger than the payments in 2001 and 2008. Most importantly, the EIPs were disbursed without any explicit randomization. Thus, while we compare our estimates to the spending responses estimated in the earlier literature, our main analysis uses an estimator that is both more robust to nonrandom differences in spending responses over time and better suited for the variation across households in the EIP programs. In terms of being more robust, our main analysis employs a method that is unbiased in the presence of significant difference in spending responses over time (for the same round of EIPs), a concern in recent literature on treatment effects (Borusyak, Jaravel, and Spiess 2021; Orchard, Ramey, and Wieland 2022).

In terms of being better suited for the variation across households in the EIP programs, each round of EIPs was distributed mostly during one month and without any random variation across months. For example, the first round of EIPs had the most variation in timing; almost half of these EIPs were disbursed by direct deposit during the week of April 10, and almost 90 percent of 2020 EIPs were disbursed within the first five weeks.[2] As a result, our main analysis leans heavily on comparing the spending of similar households that do and do not receive EIPs and that receive EIPs of different amounts relative to their typical spending amounts. Receipt

2. We do not study the spending responses to EIPs that were received as part of income tax refunds or implicitly as lower tax payments.

status is primarily driven by whether the Internal Revenue Service (IRS) had the information to disburse the payment and whether the household was ineligible due to too high income or citizenship status.[3] Section III presents our method, including how we further modify the canonical method for the extreme volatility in expenditures during the pandemic.

Our first main finding is that the CE data show only small short-term spending increases on nondurable goods and services in response to the receipt of an EIP. For the first round of payments in 2020, 95 percent confidence intervals imply that people increased their spending on nondurable goods and services as measured (roughly 44 percent of total expenditures measured in the CE) by between 4.6 and 15.8 percent of their EIP during the three-month CE reference period during which the EIP arrived.[4] We find a similar average propensity to increase consumer spending (marginal propensity to increase consumer expenditures, or MPC) for the second, smaller round of EIPs, disbursed mainly in January 2021 when the economy was somewhat more open. For the third round of EIPs in the spring of 2021, our estimates imply almost no spending response. An important caveat to these second two results is that receipt of these EIPs appears to be underreported in the CE survey, and therefore these spending responses may be underestimated. Nonetheless, all three estimated spending responses on the broad measure of nondurable goods and services in the CE survey are small and suggest that most EIP dollars were not providing urgently needed pandemic assistance.

These relatively low spending responses are consistent with the fact that the EIPs were disbursed far more broadly than the income losses caused by the pandemic, with the presence of pandemic constraints on spending, and with the large increase in household account balances during the pandemic. Roughly 145 million EIPs were disbursed by mid-2020 while employment dropped by 22 million during the pandemic recession.[5] Particularly during the first wave of EIPs, many types of consumption were constrained by the prevalence of the disease or by government restrictions which, together with

3. For the first round of EIPs for example, 3.8 percent of eligible households did not receive an EIP in 2020 because the IRS did not have the necessary information to disburse their EIP, and 16 percent of tax units were not eligible for an EIP because their incomes were too high or they did not meet the citizenship requirements, for example, a couple with one noncitizen spouse that filed jointly; see sections I and II (Murphy 2021).

4. This propensity to increase consumer spending within a few weeks of the arrival of the first round of EIPs is somewhat lower than found in previous studies using aggregated data or information on select populations, issues we discuss below.

5. Cajner and others (2020) and Cox and others (2020) document the large diversity in outcomes in the pandemic recession.

diminishing marginal utility on unaffected goods and services, could have held back the overall expenditure response to the payments. Indeed, Guerrieri and others (2022) make this assumption to study the macroeconomic consequences of the pandemic, and our results show some evidence of additional spending on durable goods for the first two EIP rounds, consistent with the shift in aggregate retail spending from services and toward durable goods during the pandemic.[6]

Particularly for the second and third round of EIPs, these low spending responses are also consistent with households on average already having plenty of liquid funds. As the constraints on spending relaxed, the pandemic reduction in spending coupled with other government support (e.g., the paycheck protection program and expanded unemployment insurance benefits), including earlier EIPs, may have raised average liquidity and lowered the need for households to spend during the second and third waves of EIPs. Finally, the third round of EIPs was large relative to all other payments, and larger transitory increases in income in theory raise liquidity themselves and lead to smaller shares of the increase being spent in the short run.

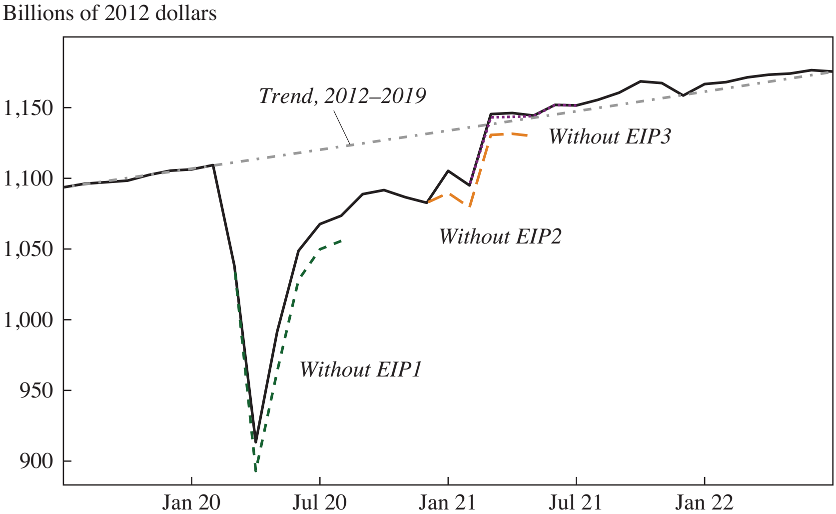
Did the EIPs cause later spending? We find some evidence of continued higher spending in the months following the three-month period of receipt, although these are fairly statistically uncertain. We estimate that the roughly 45 percent (round one) and 60 percent (round two) of people's EIPs were spent after the concurrent and subsequent three-month period. We measure essentially no spending increase in response to the third round of EIPs at any horizon. Our analysis has no power to estimate spending responses at longer horizons. However, following the disbursement of the EIPs, credit card balances decreased, liquid account balances increased, and stock prices for "meme"' retail stocks increased (see Greig, Deadman, and Sonthalia 2021; Greenwood, Laarits, and Wurgler 2022). Strong household balance sheets typically raise expenditures and so may have contributed to higher demand as the pandemic waned.

Figure 1 summarizes these findings by showing that the direct, short-run spending responses to the EIPs were relatively small. The figure plots observed real personal consumption expenditures (PCE) and the same series subtracting the increase in spending implied by our estimates assuming that the contemporaneous spending response occurs evenly over the month of

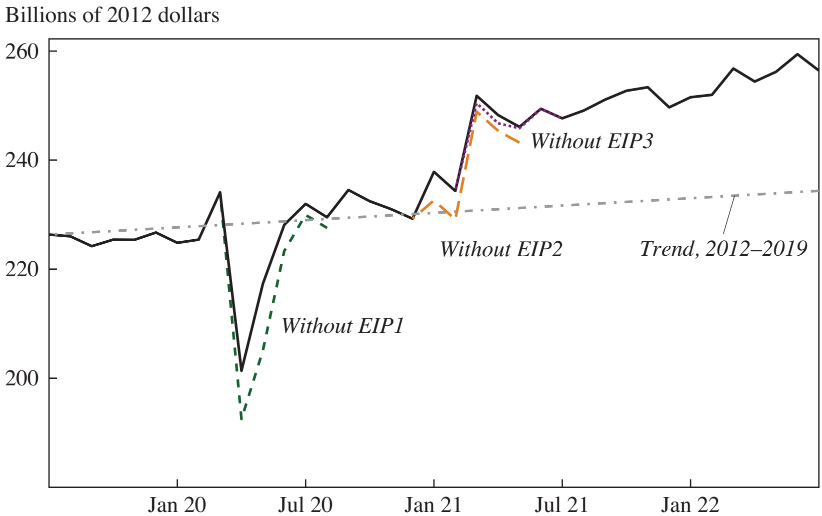
6. In total, we estimate that about 24 percent of EIPs were on average spent in the three-month period in which they arrived on all CE expenditures. The spending responses to the EIPs were on average more tilted to durable goods than the spending responses to the 2001 tax rebates but not that dissimilar from those to the 2008 economic stimulus payments.

Figure 1. Implied Change in Real Personal Consumption Expenditures Directly Due to Disbursement of EIPs

Panel A: Personal consumption expenditures



Panel B: Nondurable personal consumption expenditures



Source: Authors' calculations, based on National Income and Product Accounts.

Note: Monthly personal consumption expenditures in billions of 2012 dollars (August 17, 2022). The trend line is the average monthly growth rate of real PCE from January 2012 to December 2019 applied to the real value of PCE from July 2019. Without EIP series are constructed by subtracting from PCE the spending implied by the MPC estimates from table 4 and the monthly EIP payments from the EIP Dashboard, Bureau of the Fiscal Service, as of December 15, 2021. We assume that the contemporaneous spending occurs evenly in the month of receipt and the subsequent month, and that lagged spending occurs evenly over the following three months. We assume negative estimated spending is actually zero.

receipt and the first following month and that the lagged spending response occurs evenly over the following three months. The lines without different EIPs in figure 1 are thus not true counterfactuals but are simply PCE without the partial equilibrium effect of the EIPs on consumer spending based on this simple accounting exercise. Figure 1 not only shows the relatively small increase in direct spending implied by our estimates but also highlights the extremely strong rebound in consumer demand for nondurable goods and services to which the EIPs may have contributed with delay through temporary decreases in debt or increases in saving.[7]

Our second main finding is that while the average spending response to the EIPs is modest, we find significantly higher short-term spending responses for households that are more exposed to the economic losses from the pandemic, consistent with these households using the EIPs to fund spending that they could not easily do otherwise. Our first measure of exposure is low ex ante liquid wealth. For the first round of EIPs, households in the bottom third of the distribution of liquid wealth—those with less than \$2,000 available ex ante—spent at roughly two and a half times the rate of those in the middle third, while those in the top third of the distribution of liquid wealth (above \$12,500) had roughly no spending response. Differences in liquidity across households are less important for the second two rounds.[8] Our second measure is based on whether a household earns a significant share of its income from work that is unlikely to be able to be done from home or remotely. Households with lower ability to work from home spent more out of their first round of EIPs when they arrived. We find no such evidence for later rounds of EIPs.

In sum, while on average the EIPs appear to have gone to many households with incomes that were unharmed by the pandemic, some of the EIPs, mainly in the first round, did support short-term spending for some households, primarily those with low ex ante liquid wealth and those reliant on income that could not be earned by working from home. In terms of future policy, both this paper and the research on consumption responses to tax payments more generally suggest that greater targeting of households with little liquid wealth and low debt capacity would be more efficient

7. Note that for nondurable PCE, we use MPC estimates from a CE measure that includes some services and semidurable spending, so it likely overestimates the spending effects of the EIPs.

8. For the second round, we find essentially no spending response in the top third of the distribution of liquid wealth but similar spending responses between the bottom two thirds. Finally, the only evidence for spending in response to receipt of the third round of EIPs is for the middle third of the distribution of liquid wealth.

in the sense of generating more rapid increases in demand for purposes of stimulus programs or getting more of the payment money to those households most vulnerable to income losses for pandemic insurance.[9] However, there are also potential moral hazard costs of targeting economic need or low liquidity more directly. One approach to minimizing these costs would be to base payments on household characteristics that are less responsive, for example, not sending pandemic insurance payments to people who were not previously employed and therefore not at risk of losing their jobs (e.g., people who were retired in 2019 did not lose their jobs in 2020). Alternatively, either stimulus or pandemic insurance could be delivered through increasing temporarily the generosity or eligibility of existing government programs that are based on direct targeting, such as unemployment assistance, Temporary Assistance for Needy Families, and so on, where the disincentives of these programs are better understood and potentially better minimized (Ganong and others 2022).[10]

Most studies of the spending response to previous tax payments have estimated the response to payments using variation in spending between recipients and non-recipients (Bodkin 1959; Agarwal and Qian 2014; Kueng 2018), over time (Souleles 1999; Parker 1999; Stephens 2003; Farrell, Greig, and Hamoudi 2019; Baugh and others 2021), and using randomization in policy in either dimension (Agarwal, Liu, and Souleles 2007; Broda and Parker 2014; Parker 2017; Lewis, Melcangi, and Pilossoph 2019).[11] The disbursement of the EIPs was not randomized in any way across households or time. Because of this, the present study as well as existing studies of the spending response to the EIPs focus on comparing spending before receipt to spending after receipt, comparing spending between recipients to non-recipients, and comparing households receiving different sized EIPs.[12]

9. Past payments sent out either as pandemic insurance or as a stimulus program have increasingly targeted these populations to some extent by excluding households with high incomes the previous year.

10. For pandemic insurance, Romer and Romer (2022) also suggest a role for policy in providing hazard pay. For the purposes of economic stimulus, it is also worth noting that government spending generates immediate spending by definition, and so in this sense it is equivalent to an MPC of 100 percent out of a payment program. That is, rapid government spending raises aggregate demand by more than payment programs with equivalent costs, although obviously the goods and services purchased will differ, as will the distributional effects of the policies.

11. Most closely related, Fagereng, Holm, and Natvik (2021) measure the spending response of (random) lottery winners.

12. Kubota, Onishi, and Toyama (2021), Feldman and Heffetz (2022), and Kim, Koh, and Lyou (2020) measure the spending responses to tax payments disbursed in response to the pandemic in Japan, Israel, and South Korea, respectively.

The first rapid analysis of the spending changes caused by the EIPs, Meyer and Zhou (2020), used Bank of America transactions data and reported large increases in aggregated card spending on the day of and the day following receipt of an EIP associated with bank accounts that received EIPs on April 15 (when over 40 percent of EIPs were disbursed) relative to those that did not. Daily spending increased by an average of 50 percent year over year between April 15 and 16 for households with incomes below $50,000 and by only 3 percent for households with incomes above $125,000. Also using aggregated data, Chetty and others (2020) find that over this same couple of days, credit card spending in zip codes in the bottom quarter of the distribution of average household income rose by 25 percentage points while those in the top quarter of the distribution rose by only 8 percentage points. Finally, also using zip code–level data and using incidental differences in timing in EIP disbursements across zip codes, Misra, Singh, and Zhang (2021) infer an MPC of 50 percent in the few days after an EIP arrived.

Our evidence shows lower spending responses than measured in existing studies, all of which use account-level data on financial transactions to measure the spending. Karger and Rajan (2021), Baker and others (2020), and Cooper and Olivei (2021) find that people's out-of-account spending rises cumulatively by 46 percent, 25–40 percent, and 66 percent of their first-round EIPs, respectively, within a few weeks of receipt.[13] One likely reason for these larger spending responses than found in the CE survey data is that these account-level studies cover populations that are likely to have larger spending responses than average.[14] There are other possible reasons also, such as the different ways in which the studies measure consumer expenditures. Account-level data on transactions may mischaracterize debt payments or saving as consumption (e.g., paying debt on unlinked credit cards, payments of overdue bills from past consumption, or transfers to investment accounts).[15] Alternatively, respondents in the CE survey could

13. Karger and Rajan (2021) also estimate a 39 percent MPC for the second round of EIPs.

14. The accounts used in Karger and Rajan (2021) are skewed toward lower income households (average annual income of $20,880); the households in Baker and others (2020) are those that have opted to use a financial app designed to help them save (and have average incomes of $36,000); and Cooper and Olivei (2021) use Facteus data covering lower-income households many of whom are unbanked.

15. Baker and others (2020) include car loans and mortgage payments as consumption-related spending, whereas this paper includes interest payments on mortgage loans as part of consumption-related spending, but not payments on the principal.

forget to report EIP-induced purchases. Finally, the differences could arise in part from statistical issues, both the statistical uncertainty inherent in any estimator and the statistical methods that we use.[16]

I. The Economic Impact Payments

We organize our description of the EIP programs around the three ways in which EIPs differed across households: differences in dollar amount conditional on receipt, differences in the time of receipt of the EIP, and whether a household did or did not receive an EIP at all. Unlike when payments were disbursed in 2001 and 2008, none of these three sources of variation are completely unrelated to household characteristics.

In terms of amount, the first round of EIPs (which we call EIP1s) consisted of a base payment of $1,200 for an individual, $2,400 for a couple filing jointly, and additional payments of $500 for each qualifying dependent under age 17. The CARES Act set upper income thresholds for receiving the full payment of $75,000 for an individual, $112,500 for a head of household, and $150,000 for couples filing jointly, where income was based on 2019 adjusted gross income (AGI) if the taxpayer had already filed their 2019 tax return in 2020, otherwise income was based on 2018 AGI as reported in 2019 tax filings.[17] For every $100 of AGI over the threshold, the stimulus payment was reduced by $5.

Second-round EIPs—EIP2s—were smaller, consisting of a base payment of $600 for an individual or $1,200 for a couple filing jointly, and additional payments of $600 for each qualifying dependent under age 17. The upper income thresholds and phaseout rate for this second round of EIPs were the same as for the first round.[18]

The third round of EIPs—EIP3s—were substantially larger than EIP1s or EIP2s. They consisted of a base payment of $1,400 for an individual, $2,800 for a couple filing jointly, and additional payments of $1,400 for

16. The CE is a small data set, with a similar number of recipients to that in Baker and others (2020), and standard errors are a substantial share of the differences among the estimates across the papers. The randomness of the estimator may also explain the difference between our estimated spending propensities and those estimated in the CE during previous tax rebate episodes.

17. In December 2020, the phaseout threshold for a qualifying widow(er) increased from $75,000 to $150,000, according to the IRS. This change does not affect our analysis.

18. For the second round of EIP, income is defined as the tax filer's 2019 AGI reported on their 2020 tax filings. If a tax return had not been filed by the time the payments were distributed, the tax filer did not receive an advanced payment and had to claim the Recovery Rebate when filing their 2020 tax return in 2021.

each qualifying dependent. They were also distributed slightly more broadly along several small dimensions, and included a definition of "qualifying dependent" that was expanded to include dependents over the age of 17. The upper income thresholds were the same as in the first and second rounds; however, the phaseout rule was more aggressive so that the larger amounts did not lead to EIPs being received higher up the income distribution. Specifically, rather than a constant phaseout rate, income thresholds were set such that tax filers with a 2020 AGI above $80,000 for an individual, $120,000 for a head of household, and $160,000 for a couple filing jointly, regardless of the number of qualifying dependents, did not receive an EIP3.[19] For example, an individual with no dependents, base payment of $1,400, had a phaseout rate of $28 for every $100 of AGI over $75,000, whereas an individual with one qualifying dependent, base payment of $2,800, had a phaseout rate of $56 for every $100 of AGI over $75,000. Figure 2 displays the EIP amounts as a function of income for various family structures for the first, second, and third round of EIPs.

In addition to households receiving different amounts of EIPs, households also received them at different times. In each round, most taxpayers who had included their bank information when filing a recent tax return (e.g., for a refund) received their EIP during the first week of disbursement. For EIP1, bank information came from a 2018 or 2019 tax return, and for EIP2 and EIP3, bank information came from a 2019 or 2020 tax return. The IRS also launched a web page where households could enter their information for the IRS if they either had omitted bank information from their returns or were eligible but had not filed 2018 or 2019 returns.[20] For EIP1, this constituted roughly 30 million households (Murphy 2021). The IRS also collected information on eligible households from the Social Security Administration and the Department of Veterans Affairs (and the Railroad Retirement Board).

The IRS began depositing EIP1s into bank accounts mid-April 2020, and using the information that the IRS was able to gather and process in time, roughly 105 million or about 62 percent of all EIPs were disbursed in April 2020 (Murphy 2021). For eligible households without the necessary bank information, the EIPs arrived starting in mid-April by mailing a paper check or prepaid EIP card. The disbursement of checks occurred

19. If a 2020 tax return had not yet been filed, then 2019 AGI from the 2019 tax return filed in 2020 was used instead.

20. IRS, "Get My Payment," https://www.irs.gov/coronavirus/get-my-payment; no longer available, but as of this writing, there are links to further information.

Figure 2. Economic Impact Payment Amounts as a Function of AGI and Family Structure

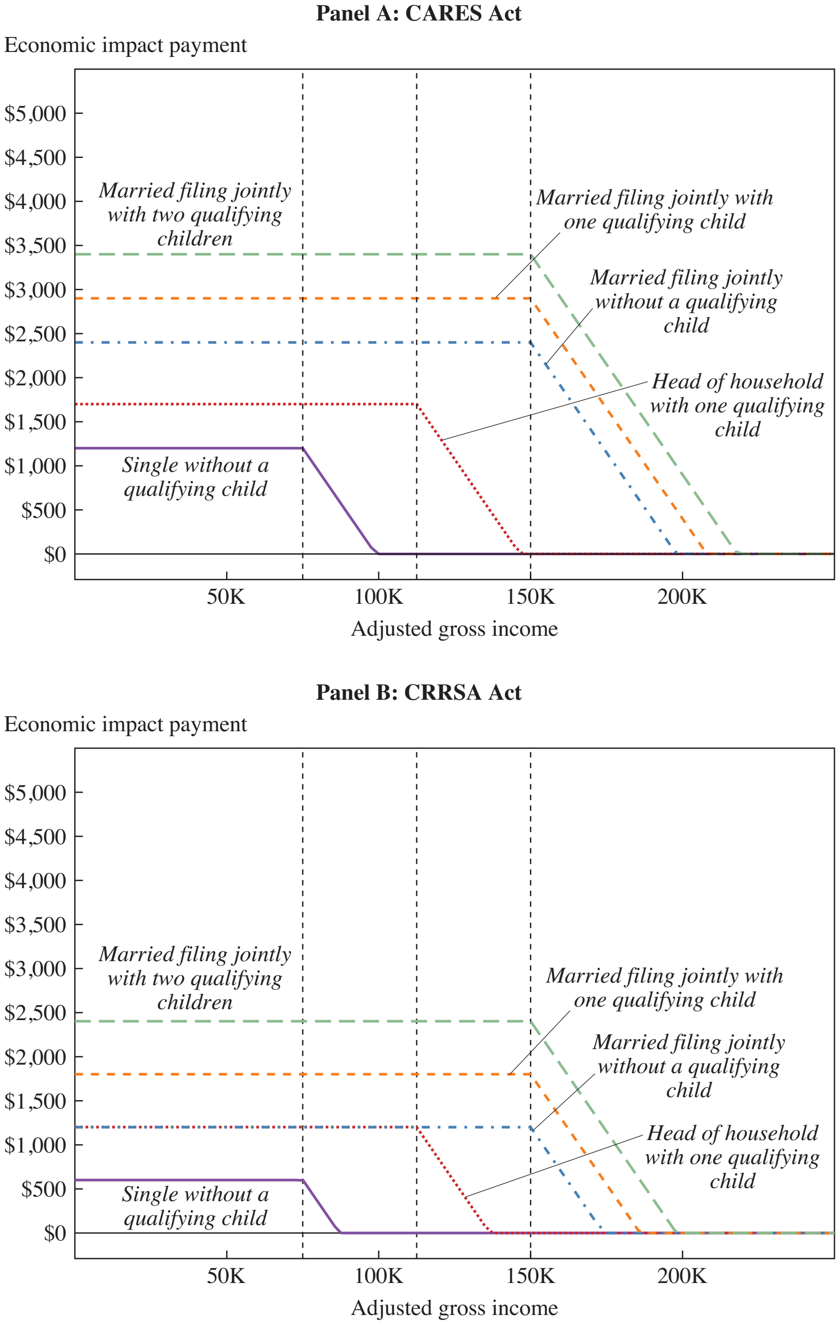
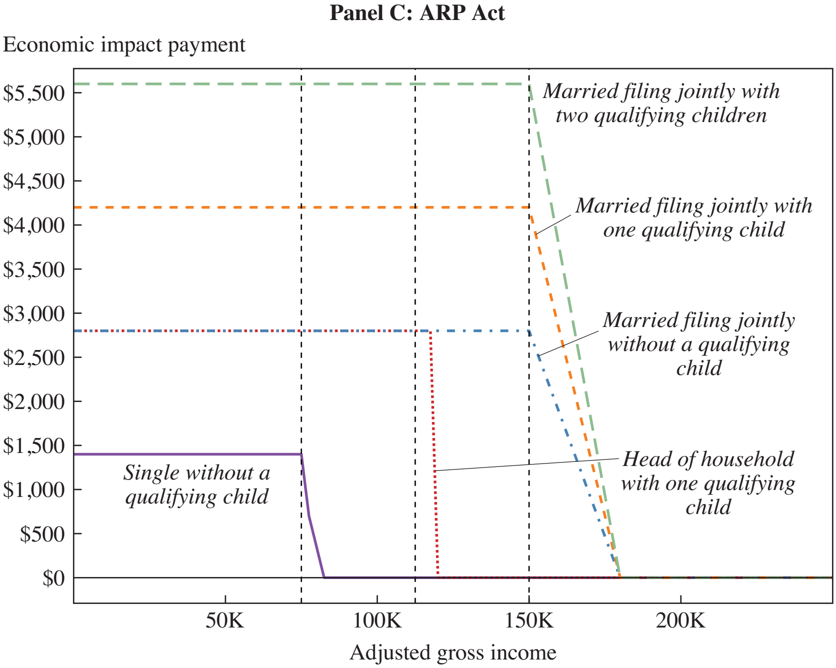


Figure 2. Economic Impact Payment Amounts as a Function of AGI and Family Structure (*Continued*)



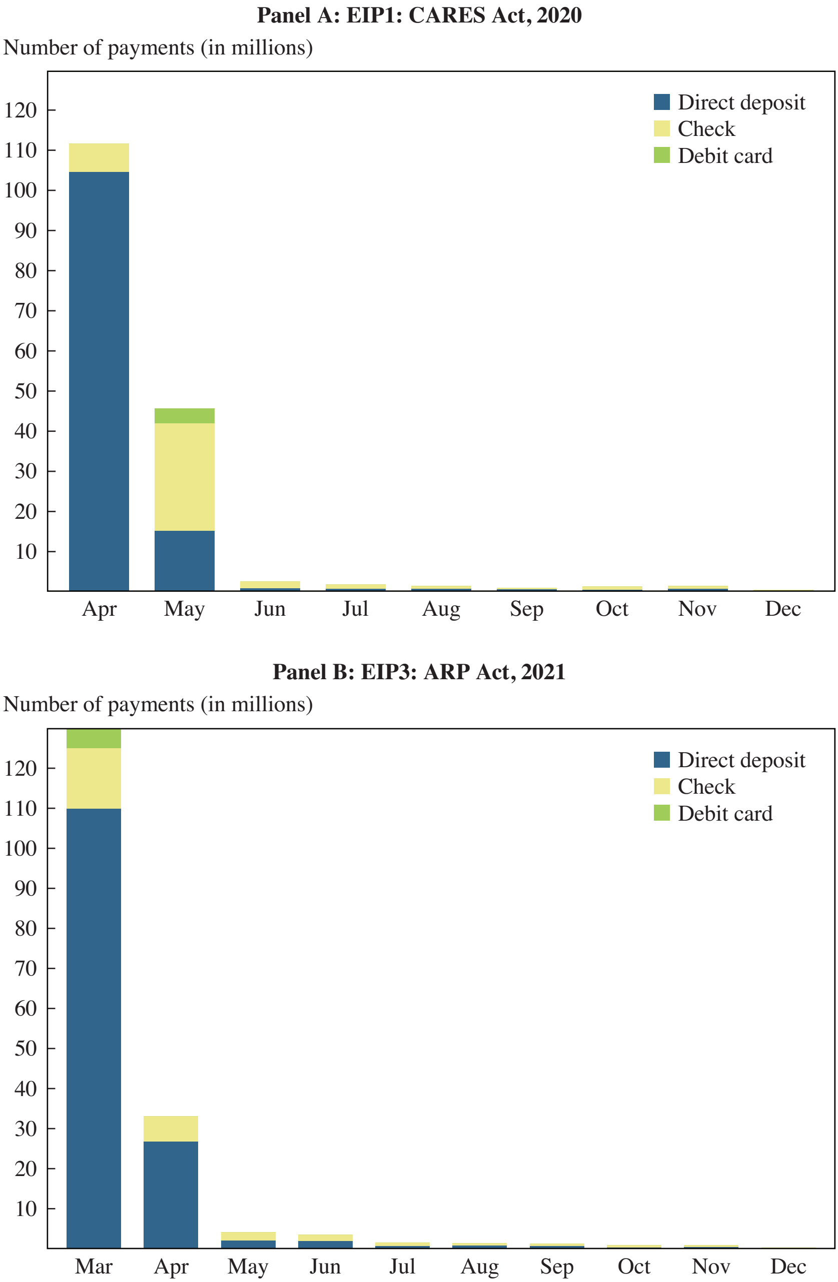
Source: Authors' calculations.

with a greater delay. By the end of April only about 7 million checks (4 percent of EIPs) were sent out. Most of the checks were sent out in May, about 27 million or 16 percent of EIPs, and all of the EIP cards were sent out in May, about 4 million or 2 percent of EIPs. About 95 percent of all first round EIPs were delivered in the first two months of disbursement. The remainder of the EIPs continued to trickle out over the rest of 2020. Figure 3, panel A, shows the minimal variation in timing of the distribution of CARES Act EIPs.

In contrast, the disbursement of the second round of EIPs has almost no variation across months. Almost all of the EIP2s were distributed in January 2021. Daily Treasury statements show some EIP2s were also being disbursed in February, which was due to reissuing payments that were initially unable to be delivered.

The disbursement of the third round of EIPs was slightly more drawn out over time than that of the EIP2s, but still more concentrated over time

Figure 3. The Disbursement of EIP Payments over Time and by Mode of Distribution



Source: Bureau of the Fiscal Service.
Note: Months are the disbursement months.

than the first round of EIPs. A full 74 percent of all EIP3s were distributed in March 2021 (62 percent by direct deposit; 8.5 percent by check; and 2.7 percent by EIP cards). By the end of April about 92 percent of all third-round EIPs had been distributed, with the remaining 8 percent distributed over the remainder of 2021. Although the IRS distributed a smaller percentage of EIP3 in the first two months of disbursement compared to EIP1, about 5 million more EIPs were distributed during March and April of 2021 than compared to April and May of 2020. Additionally, about 20 million (7 percent) more EIPs were distributed by direct deposit. Figure 3 displays the variation in the timing of disbursement of EIP1s and EIP3s.

Finally, there is a set of households that either did not receive EIPs at all or who received their EIPs after filing their taxes as part of their income tax refunds or implicitly as reduced income tax payments. There are three main reasons why a household did not receive an EIP during each primary disbursement period. First, an individual was ineligible for an EIP if they did not have a Social Security number (SSN) valid for employment. The CARES Act was worded such that families were ineligible if they had filed jointly and one of the spouses was not a US citizen, a situation affecting an estimated 14.4 million people (Gelatt, Capps, and Fix 2021). The CRRSA Act changed this requirement. A married couple filing a joint return became eligible for a partial recovery rebate credit when only one spouse has an SSN. This change resulted in 2.9 million people becoming eligible.[21] The ARP Act further expanded the eligibility criteria to anyone with an SSN, which resulted in an additional 2.2 million eligible individuals.[22]

Second, eligible households did not receive an EIP disbursement if they had changed accounts or addresses during the relevant previous year, if they had not given their information to the IRS, or if the IRS did not otherwise have their information (e.g., from the Social Security Administration). For example, four months after the CARES Act, 10 percent of EIPs had not been disbursed, and 5 percent or 9 million eligible households had not received an EIP by the end of September (Murphy 2021). For EIP2 or EIP3, people who relocated even temporarily during the pandemic and

21. Of these 2.9 million people, 1.4 million were US citizens or legal immigrants and spouses of an unauthorized immigrant, and 1.5 million were children with one unauthorized immigrant parent. The change in eligibility criteria was applied retroactively, which means not only did these individuals now qualify for the second EIP, but they were also able to claim the first EIP through the recovery rebate tax credit on their 2020 tax filing.

22. These 2.2 million individuals are children whose parents (or parent) are unauthorized immigrants. Since no parent had an SSN, they were ineligible for the first and second EIPs, which means their children were also ineligible.

formally changed their addresses or banks accounts became ineligible for EIP disbursement.[23]

Finally, the third reason for households not receiving an EIP is that EIP amounts declined to zero as income increases. As shown in figure 2 high-income households were not eligible, and a significant number of higher-income households that received EIPs in the first two rounds were not eligible for an EIP3.

Taxpayers who fell into either of the first two categories and so did not receive a disbursed EIP but were eligible for an EIP could receive their EIPs as tax credits when they filed their 2020 taxes in 2021 for EIP1 and EIP2, and when they filed their 2021 taxes in 2022 for EIP3. More generally, taxpayers were also eligible to receive a tax credit for any amount by which the EIP they were due based on their final tax information exceeded the amount they had been disbursed. These true-ups amounted to roughly \$45 billion in tax year 2020 and \$18 billion in tax year 2021 (Splinter 2022). There was no corresponding payment required, however, if a disbursed EIP exceeded the amount that should have been disbursed based on the later tax information.[24]

In aggregate \$271 billion was disbursed during the first EIP round, \$141 billion during the second EIP round, and \$402 billion during the third EIP round.[25] Alone, any one of these rounds is much larger than the previous 2008 program which disbursed \$120 billion in 2020 dollars, which in turn was close to double the total of the 2001 rebate program. Combined, the three rounds of EIP disbursed more than six times the amount disbursed with the 2008 program. About \$260 billion worth of EIPs were disbursed in the second quarter of 2020, which corresponds to about 5.2 percent of GDP or 7.9 percent of PCE in that quarter (figure 3; IRS 2020). The first quarter of 2021 saw \$473 billion of EIPs disbursed, from both the second and third rounds. This represents 8.7 percent of GDP and 3.2 percent of PCE during the quarter. The third EIP round additionally disbursed \$67 billion in the second quarter of 2021, corresponding to 1.2 percent of GDP and 1.8 percent of PCE. The next section describes the EIPs as recorded in our CE data set.

23. IRS, "2021 Recovery Rebate Credit Questions and Answers," https://www.irs.gov/newsroom/2021-recovery-rebate-credit-questions-and-answers.

24. These safe harbor amounts were roughly \$21 billion in tax year 2020 and \$44 billion in tax year 2021 (Splinter 2022).

25. IRS, "SOI Tax Stats—Coronavirus Aid, Relief, and Economic Security Act (CARES Act) Statistics," https://www.irs.gov/statistics/soi-tax-stats-coronavirus-aid-relief-and-economic-security-act-cares-act-statistics.

II. The Consumer Expenditure Survey

Data for this study are from the Consumer Expenditure (CE) Interview Survey, a household survey run by the Bureau of Labor Statistics (BLS).[26] The CE data set contains spending, demographics, and other financial information on households living in the United States. The BLS structures the CE so that a consumer unit (CU) at a given address, which we will refer to as a household, is interviewed up to four times at three-month intervals about their spending over the previous three months (the reference period). New CUs are added to the survey every month, and while a significant dollar share of spending data is reported at the monthly level, a little over half of spending is only reported for the entire three-month reference period. Thus, we use the data at the (overlapping) three-month frequency.[27] Online appendix A.2 contains more details about CE files and variables we use in this study.

Following the passage of the CARES Act, the BLS added a module of questions about the EIPs to the CE survey, starting with the June 2020 interviews and continuing until the October 2021 interviews, with the exception that the questions were not fielded in January 2021.[28] These questions were worded similarly to questions that the BLS added to the CE about stimulus payments in 2008. The questions measure the date of receipt, the number of EIPs received, the amount received, which member or members of the household the payment was for, and the mode of receipt (by check, direct deposit, or debit card).[29] The questions were phrased to be consistent with the style of other CE questions and the questions on previous CE surveys about the 2001 and 2008 tax rebates. Although the wording did not exactly follow previous CE surveys, the

26. Information on the data and methods of survey can be found at US Bureau of Labor Statistics, "Consumer Expenditure Surveys," https://www.bls.gov/cex/.

27. "Overlapping" means that for CUs interviewed within two months of each other the three-month reference period for reporting spending will include some of the same months. For example, a CU who is interviewed in June has a three-month reference period of March, April, and May, and a CU interviewed in July has a three-month reference period of April, May, and June. Both reference periods include April and May; thus, we consider them overlapping.

28. The module was developed by the BLS partly based on the similar questions from 2008 and in consultation with others in the federal statistical system, particularly those working with the Household Pulse Survey (in which EIP questions had already been asked) and outside researchers, two of whom are coauthors of this paper.

29. Starting with the July interview the mode of receipt question was expanded to include via tax rebate. Any instances of receipt via tax rebate were dropped, which resulted in five relevant rebates being excluded. Prior to the July interview, CUs who received an EIP via their tax rebate were asked to not include it when reporting EIP receipt.

module of questions also asked whether the EIP was used mostly to add to savings, mostly to pay for expenses, or mostly to pay off debt. Online appendix A.1 contains the language of the CE survey instruments.

The fact that the EIP questions were not included in the May 2020 interview questionnaire means that, even for EIP1 where the distribution of EIPs was somewhat drawn out over time, we have very little power to identify the impact of the arrival of EIP1s on spending using only variation in the timing of receipt across households. The vast majority of EIP1s were disbursed in April and May. And while April and May are in different three-month expenditure recall periods for households on the May interview cycle, they are not for households on the June or July interview cycle. Thus, we cannot compare how spending differs between April and May depending on whether an EIP1 is received in April or May. Since EIP2 and EIP3 have very little variation in the timing of receipt, and since only about 10 percent of EIP1s arrive after May 2020, we focus primarily on analysis that leans heavily on other sources of variation, like amount and recipient status.[30]

A second implication of the lack of EIP questions on the May 2020 survey is that we have no way to tell whether households interviewed in May received the EIP1 or not during the previous three months. The reference period for the May 2020 interview includes April, when over half of all EIPs were disbursed. Thus, we drop all households on this interview cycle because we cannot compare the spending of those receiving different EIPs at different times (or not at all) since we do not have the EIP information. More precisely, we restrict our sample to households that had an interview during June or July of 2020 when the EIP questions were asked and the three-month recall periods include April and May 2020. This restriction drops roughly one-third of households—those in the interview cycle that includes May 2020, as well as any other households that are missing interviews in June or July 2020 interviews. To be clear, we use all available interviews for the households that have interviews in June or July 2020 (provided the observation has the other necessary data and a consecutive interview also with valid data). However, the loss of the observations on the May interview cycle reduces statistical power.

We face a similar but less significant challenge for households interviewed in January 2021. In this case, we assume no EIPs were received in

30. We investigated measuring the spending response to the EIP1 using the data at the monthly frequency and only the CE expenditure categories that are collected by month but found weak statistical power (consistent with the conclusions of prior work with the CE).

the reference period (October, November, and December) for households interviewed in January 2021 (when the EIP questions were not asked).[31]

We construct two main samples of CE households for each EIP round. For each round, we limit the sample to households with interviews during the main period of disbursement: June and July 2020 for EIP1, February, March, or April 2021 for EIP2, and April, May, or June for EIP3. For each, we construct first a broad sample we refer to as all households that makes minimal additional drops and follows exactly earlier analyses of tax rebates in the CE. Second, motivated both by the unprecedented nature of the pandemic and programmatic differences between the EIPs and previous tax rebates, we construct our final sample by adjusting the way in which older households and households with very low levels of reported expenditures are dropped and dropping high-income households who are mostly ineligible for EIPs. (See details in online appendix A.5.3 and table C.5–C.7.) We discuss these choices in detail in the next section.

Table 1 shows that the monthly distribution of EIPs reported in the CE line up reasonably well with other official statistics. The first two columns of table 1 show statistics for our final sample (which drops high-income households as described subsequently); the second two columns show statistics for the CE data including all (available) interview months. For EIP1, April data for the raw CE sample are adjusted up by 50 percent to account for our dropping one-third of recipients, those interviewed in May when the EIP questions were not asked. The CE data have slightly fewer EIP1s reported during the peak month of April and more in the following months than the US Treasury reports. This difference is consistent with some time delay between disbursement and receipt for mailed payments and with some households taking time to notice EIPs deposited into their accounts (and with the possibility that some households report a later date of receipt than actually occurred).[32] For later rounds of EIP, the monthly distribution lines up well with what we know from other sources also.

Columns 3 and 4 of panel B in table 1 show that 24 percent of households do not receive an EIP1 according to the CE data compared to 20 percent in reality (3.2 percent of households were eligible tax units who were

31. Less than 2 percent of EIP1s were distributed over October, November, and December. EIP2s began being distributed by direct deposit during the last few days of December but did not clear until January 4, the official payment date according to the IRS. Checks for EIP3 did not begin being distributed until January (Murphy 2021).

32. In the final sample, about 10 percent of households that get EIPs report multiple EIPs. About 50 percent of these report EIPs in more than one month of which about 60 percent report receiving EIPs in different reference periods.

Table 1. Percentage of EIPs by Month and Percentage of Households Not Receiving EIPs

	Unweighted CE final sample (1)	*Weighted CE final sample (2)*	*Unweighted CE (adjusted) (3)*	*Weighted CE (adjusted) (4)*	*Census Bureau's Household Pulse Survey and US Treasury (5)*
Panel A: The distribution of EIP1s across months, as a percentage					
April 2020	53.8	54.6	53.1	54.1	66.4
May 2020	36.3	35.4	35.3	34.3	25.7
June 2020	7.5	7.7	8.9	9.0	1.1
July to November 2020	2.4	2.3	2.7	2.6	6.8
Panel B: Percentage of households or tax units not receiving an EIP1					
Total (households)	17.0	17.0	24.7	24.6	
Ineligible (tax units)					16.2
Eligible (tax units)					3.2
Panel C: The distribution of EIP2s across months, as a percentage					
December 2020	24.3	24.2	19.6	19.4	0
January 2021	68.6	68.5	64.2	63.7	100
February 2021	7.1	7.3	16.2	16.9	0
Panel D: Percentage of households or tax units not receiving an EIP2					
Total (households)	50.8	51.9	52.2	53.0	
Panel E: The distribution of EIP3s across months, as a percentage					
March 2021	68.2	68.8	65.8	66.2	73.8
April 2021	23.7	23.3	25.9	25.8	18.8
May 2021	3.5	3.2	3.6	3.4	2.3
June to December 2021	4.6	4.7	4.6	4.6	5.2
Panel F: Percentage of households or tax units not receiving an EIP3					
Total (households)	29.4	29.0	40.5	40.3	

Source: Authors' calculations.

Note: Weighted data using the average of FINLWT21 across all interviews. All samples use available CE data, so interviews through and including September 2021. See online appendix A.5.3 for CE sample construction and adjustments for months in which EIP questions were not asked. "Unweighted CE" includes all households with interviews in these months. In panels A, C, and E, months are recipient months in the first four columns but are disbursement months in the last column. In the final column of panel B ineligible households is as self-reported in the Census Pulse Survey from Garner, Safir, and Schild (2020), and eligible households not receiving payments are counted through October 2020 as reported in Murphy (2021). For panels C and E, the disbursement data come from the Bureau of the Fiscal Service, US Department of the Treasury.

non-recipients in 2020, and 16 percent of households were not eligible for EIPs). In our final CE data set, about 17 percent of households do not receive an EIP1 because we drop households with high incomes. As shown in panels D and F, these numbers are larger for EIP2 and EIP3, and while EIP3 was phased out more rapidly with income, so that fewer households received the payments, these numbers suggest that the CE data are missing some EIPs.

In terms of dollar amounts, the average value of EIP1s received in a reference period, conditional on a positive value, is $2,098, slightly higher than the average individual EIP of $1,676 reported by the IRS.[33] The average EIP2 amount is $1,301, and the average EIP3 amount is more than double this amount, $2,814. Online appendix tables C.1, C.2, and C.3 show the distribution of total EIP amounts received across household reference periods in our CE final sample (unweighted, unadjusted) and that households (correctly) report most amounts at the standard EIP amounts disbursed in each round. For example, consistent with the payments specified by CARES, most reported EIP1s are at the base amounts or in multiples of $500 above them: about 55 percent of households report payments of $1,200 (the basic payment for a single filer) or $2,400 (a couple filing separately or getting the basic payment as joint filers).

According to the IRS, there were 162 million first-round EIPs disbursed in 2020 totaling $271 billion, 147 million second-round EIPs totaling $141 billion (as of early February 2021), and 168 million third-round EIPs disbursed in 2021 totaling $402 billion.[34] In the weighted CE data, and scaling up for the interviews missing for first-round EIPs, we find 138 million first-round EIPs totaling $261 billion, 79 million second-round EIPs totaling $106 billion, and 111 million third-round EIPs totaling $254 billion.[35] Households that receive EIP1 and EIP2 by direct deposit on average have slightly higher expenditures, are slightly younger, have higher incomes, have lower liquidity, and have larger EIPs than households that receive

33. IRS, "SOI Tax Stats—Coronavirus Aid, Relief, and Economic Security Act (CARES Act) Statistics," https://www.irs.gov/statistics/soi-tax-stats-coronavirus-aid-relief-and-economic-security-act-cares-act-statistics. When using all CE data, and without aggregating to the three-month reference period level, the average (unweighted, unadjusted) EIP is $1,837.

34. IRS, "SOI Tax Stats—Coronavirus Aid, Relief, and Economic Security Act (CARES Act) Statistics," https://www.irs.gov/statistics/soi-tax-stats-coronavirus-aid-relief-and-economic-security-act-cares-act-statistics.

35. The lower number in the CE for first-round EIPs is in small part a result of not including information from CE interviews after December 2020, and similarly for third-round EIPs, since data from interviews after September 2021 are not yet published.

Table 2. The Share of EIPs by Method of Disbursement and Reported Main Use

	EIP1	*EIP2*	*EIP3*
Panel A: Distribution of payment methods, as a percentage			
By direct deposit	74.5	77.7	84.6
By check	23.4	15.8	11.7
By debit card	2.1	6.5	3.7
Panel B: Distribution of reported main use, as a percentage			
Mostly for expenses	56.4	54.5	51.9
Mostly paid off debts	17.8	19.8	19.1
Mostly added to savings	25.9	25.7	29.0

Source: Authors' calculations.

Note: Statistics based on CE final sample include only CE households with certain interviews (June or July 2020 for EIP1; February, March, or April 2021 for EIP2; and April, May, or June 2021 for EIP3), with income that does not exceed a certain threshold determined by marital status and family structure, and with cleaning as described in online appendix A.5.3. Weights applied are the average of CU weights across all interviews for that CU.

the payments by mail; but for EIP3, households that receive the payments by direct deposit are slightly older and have lower incomes.

The fractions of EIPs reported by households as received by direct deposit, by paper check, and by debit card match very closely the fractions reported by the Treasury as disbursed by these methods. Panel A of table 2 shows that 75 percent of EIP1s in the CE were reported as being received by direct deposit, 23 percent by paper check, and 2 percent by debit card. The Treasury reports that 76 percent of EIP1s were disbursed by electronic deposit, 22 percent by paper check, and 2 percent by debit card during 2020.[36] Though there were no explicit instructions, CE respondents likely reported EIPs that were deposited onto federal benefit cards (Direct Express cards) as received by debit card, and while directly comparable numbers from the Treasury are not available, through June 2020, 3 percent of EIP1s had been distributed by debit card and an additional 1 percent by deposit onto benefit cards (Murphy 2021). Consistent with the increase in direct deposit across rounds, the CE shows the share of households receiving their EIP by direct deposit increasing in each subsequent round.

The BLS also asked households to report on the CE survey whether they spent or saved their EIPs.[37] The responses suggest greater spending than our analysis of expenditures does. Panel B of table 2 shows that 56 percent

36. IRS, "SOI Tax Stats—Coronavirus Aid, Relief, and Economic Security Act (CARES Act) Statistics," https://www.irs.gov/statistics/soi-tax-stats-coronavirus-aid-relief-and-economic-security-act-cares-act-statistics.

37. This is the reported preference methodology of Shapiro and Slemrod (1995).

of households report using their EIP1s mostly for expenses, and this fraction declines slightly across EIP rounds. There is also a significant increase in the share of households reporting mostly saving their EIPs in round three relative to earlier EIPs. In 2008, the BLS added different questions to the CE survey that were more similar to those in Shapiro and Slemrod (1995, 2009) and found that 32 percent of households would "mostly spend" their tax payments and 51 percent would "mostly pay down debt."

More comparable over time, Sahm, Shapiro, and Slemrod (2010, 2020) ask the same questions in both 2008 and 2020 (not in the CE survey) and the changes in answers suggest only very slightly lower spending responses in 2020 than in 2008. In response to the EIPs, 18 percent of respondents report that their EIPs will cause them to "mostly increase spending," only 1 percent lower than in 2008, which suggests little difference in rate of spending between the EIPs and earlier stimulus payments.[38]

Following previous research on spending responses using the CE, we construct four measures of consumer expenditures at the three-month frequency: (1) food, which includes food consumed away from home, food consumed at home, and purchases of alcoholic beverages; (2) strictly nondurable expenditures, which includes some services and adds expenditures such as household operations, gas, and personal care following Lusardi (1996); (3) nondurable expenditures on goods and services, which adds semidurable categories like apparel, reading materials, and health care (only out-of-pocket spending by the household) following previous research using the CE survey; and (4) total expenditures, which adds durable expenditures such as home furnishings, entertainment equipment, and auto purchases.[39]

Relative to the administrative data used in the studies of the EIPs discussed in the introduction, there are three main advantages of using the CE interview survey as well as three weaknesses. The first advantage is that the CE contains detailed measures of consumer expenditures rather than just the transaction counterpart or, for some transactions like checks

38. Garner and Schild (2021), Garner, Safir, and Schild (2020), and Boutros (2021) provide in-depth analysis of the US Census Bureau's Household Pulse Survey in which 59 percent of respondents state that they "will mostly pay for expenses" with their EIPs. More similar to Sahm, Shapiro, and Slemrod (2020), Coibion, Gorodnichenko, and Weber (2020) show that only 15 percent of households in the Nielsen consumer panel report that they mostly spent or expect to spend their EIPs. Among these households, the average spending rate is 40 percent. Armantier and others (2020) report a slightly larger number in the Federal Reserve Bank of New York Survey of Consumer Expectations in which households on average say that they consumed 29 percent of their EIPs.

39. The exact definitions are given in online appendix A.3.

or cash, just the amount.[40] Second, the CE tracks spending and EIP receipt by individual consumer units, rather than by accounts (and linked credit or debit cards). Finally, the CE is a stratified random sample of US households constructed by the US Census and so when weighted is representative of the US population (along the dimensions of the Census-based strata and conditional on participation in the survey). The main weaknesses relative to existing studies are the relatively small sample size, sampling (e.g., nonresponse) error, and the presence of measurement error in expenditures and EIP receipt.

The next section discusses how and why our estimation methodology differs from previous approaches, as well as presenting the results of applying the previous methodology exactly to estimate the average spending response to the EIPs. The following section presents our baseline estimates of spending rates based on an approach that accounts for the differences both between previous tax rebates and the 2020 EIPs, and between previous recessions and the pandemic recession.

III. Estimation Method

In this section, we first briefly present the way Johnson, Parker, and Souleles (2006) and Parker and others (2013) estimate the consumer expenditure responses to the tax rebates disbursed in 2001 and 2008. We then refine this methodology and adopt identifying assumptions that are better suited to estimating the spending effects of these EIPs given programmatic differences, the pandemic situation, and the possibility of cross-cohort differences in spending propensities within each EIP round.

Using samples analogous to our sample of all CE households, the previous papers estimate an equation analogous to the following equation for household i with consumer expenditures $C_{i,t}$ observed for (overlapping) three-month period t:

$$\Delta C_{i,t} = \sum_{s=0}^{S} \beta_s EIPn_{i,t-s} + X_{i,t}\gamma + \tau_t + \epsilon_{i,t}. \tag{1}$$

The key regressor is $EIPn_{i,t-s}$, the total dollar amount of economic impact payments from round $n \in \{1, 2, 3\}$ received by household i in three-month

40. For example, terms like “Amazon” or “Starbucks” or “Sammy White’s.” Payments to unlinked credit cards and transfers to other accounts are also difficult to categorize as spending for consumption, debt payment, or saving.

period $t - s$.[41] The variable τ_t is a complete set of time effects for every period in the sample that control for the seasonal variation in consumer expenditures as well as the average effect of all other concurrent aggregate factors. The control variables $X_{i,t}$ contain age ($age_{i,t}$) and change in family size ($\Delta FamSize_{i,t}$) which control for the life-cycle pattern of spending and for changes in consumption needs. Finally, ϵ captures movements in consumer expenditures due to individual-level factors such as changes in income, expectations, and consumption needs, as well as measurement and recall error in expenditures.

Provided ϵ is uncorrelated with the other right-hand-side regressors (and for now maintaining the assumption that β—or its distribution over i—does vary with EIP arrival date), the key coefficient β_s measures the average partial equilibrium response of household consumer expenditures to the arrival of the EIP during the three-month period s periods after the EIP arrives. In our main analysis, in which $EIPn_{i,t-s}$ is regressed on ΔC, β_s measures the share of the EIP spent, or the marginal propensity to increase consumer expenditures (MPC).[42] These estimated MPCs are based on three sources of variation: whether a household receives an EIP or not, variation in the (overlapping) three-month period in which the EIP is received, and variation in the amount of the EIP.

As we show at the end of section IV, estimates of the spending responses based on this exact methodology—while having the advantage of being most comparable to earlier work—are small, statistically weak, and unstable compared to these earlier analyses. The first finding may simply reflect reality, but the second two may be indicative of problems with the methodology, driven by differences between previous tax rebate programs and this one, differences between previous recessions and the pandemic recession, and concerns raised recently about consistent estimation if MPCs vary across households such that the distribution of $\beta_{s,i}$ changes over time.

Our first main concern is differences between previous tax rebates and these EIPs. Relative to the earlier studies, the timing of the disbursement of the EIPs was not randomized in any way and was far more limited, both in reality (as described in section I) and observed in our data (for the reasons

41. In table 3 and in additional results in the online appendix, we sometimes replace this regressor with $\mathbb{1}[EIPn_{i,t-s} > 0]$, an indicator variable for whether an EIP from round n is received (in the period $t - s$) at all. In the online appendix, we present some results that use change in log consumption as the dependent variable.

42. When $\mathbb{1}[EIPn_{i,t-s} > 0]$ is regressed on ΔC, β_s measures the dollars spent. And when $\mathbb{1}[EIPn_{i,t-s} > 0]$ is regressed on ΔlnC, $100 * \beta_s$ measures the percentage increase in spending.

described in section II). Therefore our estimation necessarily relies more on differences in spending patterns across households with different EIP amounts, including those that do not receive EIPs (at least only as part of lower tax payments or higher refunds in the following year).

Our solution is to make the sample of non-recipients more similar to recipients by excluding households with high incomes from our analysis. Motivated by the phaseout of the EIPs described in section I, for each EIP round, we first posit an income cutoff at the nearest $25,000 above the income level (rounded to the nearest $25,000) at which a household would no longer receive an EIP. Different cutoffs apply to households with different family structures—whether the household contains children and whether it has one single adult, a married individual or couple, or multiple adults. In addition, note that recipient status is not a clean function of CE income because EIPs are disbursed based on adjusted gross income rather than the pretax income we observe in the CE, because reported income has some error, and because the IRS uses calendar year income for either 2018 or 2019 and neither year nor filing status is collected as part of the CE survey.[43] Thus, we adjust each income cutoff up in increments of $25,000 until more than 80 percent of the observations with incomes in the $25,000 range just above the cutoff are from non-recipients. Additionally, we set the cutoff for households with children to be no lower than the cutoff for households that are otherwise the same but without children (i.e., married without children and married with children), if the former has a lower cutoff after increments.[44] This process omits a few recipients. However, more importantly, it leaves some households in our analysis who are non-recipients due to having too much income but who still have incomes similar to our recipients and who therefore are potentially a good comparison group for those households who do receive EIPs. We refer to the three resulting samples—one for each EIP round—as our final samples, and it is these samples that are tabulated in section II.

Another difference between previous tax rebates and these EIPs is that there are three waves of EIPs in reasonably rapid succession, and in

43. Information on income is collected as part of the CE survey, but these questions ask about income earned in the past twelve months, which may not correspond to a calendar year. Additionally, tax filing status is not asked about in the survey, but imputed values are provided in the data. Imputations of filing status and tax liabilities are done using the National Bureau of Economic Research's TAXSIM program.

44. Online appendix tables C.5, C.6, and C.7 show the selection of resulting cutoffs and the number of recipients in the $25,000 income ranges above and below each cutoff.

equation (1), the estimated spending responses to one EIP may be biased by responses to other EIPs. In response, in our main analysis of the spending responses to EIP2 and EIP3, we include in X as control variables the same distributed lags of the other two EIPs when observed as we do for the main EIP of interest. This control is imperfect since we do not observe all earlier EIPs received and since there is cross-household correlation between recipient status and potentially even amount for EIP2 and EIP3. Thus we also check (and find similar results) when we estimate our responses without these controls.

Our second main concern is related to the fact that the pandemic was a time of unprecedented consumption volatility during which people with different levels of consumer expenditures had vastly different dollar changes over time. During the early stages of the pandemic in particular, households with higher incomes have much larger changes in dollar spending on average.[45] These differences across households suggest that the time dummies in equation (1) do a poor job of capturing the average dollar change in spending for households with different incomes. Since income and average expenditure are also related to recipient status and EIP amount, these differences may well create bias in estimates of MPCs. For example, if there are large changes in dollar spending in April 2020, when most EIP1s were disbursed, that are not caused by EIP receipt or amount conditional on receipt and yet correlated with receipt or amount, then estimates from equation (1) would be inconsistent.[46]

However, groups of people with different incomes—and so with different average levels of consumption spending—experienced roughly similar percentage changes in consumer spending over time (Cox and others 2020). We find, for example, that for a given time period t, differences in ΔlnC across terciles of the income distribution are lower than differences in ΔC (see online appendix figure C.1, panels b and c).

Our solution therefore is to scale all the variables in our regression by $\overline{C}_i$, the average consumer expenditure (of each type) for family i and also to allow a different regression intercept for households that never receive a given EIP. Letting $\tilde{X}_{i,t} = X_{i,t}/\overline{C}$ for any variable X and $R(i)$ be an indicator

45. Online appendix figure C.1, panels a and c, show this across terciles of the income distribution.

46. Previous recessions analyzed in earlier work had less variation in average change in dollar spending by income, and previous analyses found similar MPCs across different specifications, most importantly between results using log change in consumer spending and those using dollar change.

variable that equals one for households that receive at least one EIPn, we infer MPCs from the equation:

$$(2) \qquad \Delta\tilde{C}_{i,t} = \sum_{s=0}^{S} \beta_s \widetilde{EIPn}_{i,t-s} + \tilde{X}_{i,t}\gamma + \tau_t + \alpha_{R(i)} + \epsilon_{i,t}.$$

where X contains (scaled) age, change in family size, and possible previous EIPs. The main coefficient of interest, β_s, still measures the propensity to spend out of an EIP, but by scaling all variables we have transformed the τ from absorbing the average change in dollar spending across households in that period to absorbing the average percentage change in consumer expenditures across households in that period. Similarly, $\alpha_{R(i)}$ allows a different average growth rate of expenditure between recipients and non-recipients, and the residual is in terms of a percentage deviation of consumer expenditure rather than dollar deviation. In the CE survey, the average percentage change in spending measured in this way is significantly more similar for households across terciles of standards of living as measured by their average level of income (compare online appendix figure C.2, panel a to panel b and panel c to panel d).

Our third and final main concern is related to the developing literature addressing potential bias in difference-in-differences type estimation with both different groups treated at different times and heterogeneity in average treatment effect across groups (Borusyak, Jaravel, and Speiss 2021; de Chaisemartin and D'Haultfuille 2020; Goodman-Bacon 2021; Sun and Abraham 2021; Callaway and Sant'Anna 2021; Wooldridge 2021). In our context, estimation of equation (1) would be biased if there is variation in average MPC, or β_s, across households receiving the EIP in question in different months. The bias would arise from (implicitly) comparing the expenditure responses of households receiving EIPs at different times to infer the evolution of expenditure after EIP receipt. Equation (1) assumes that each household's expenditure response is given by β_s instead of $\beta_{s,t}$.[47] To be clear, any variation in the tendency to spend out of EIPs in different rounds (one, two, and three) would not create any bias.

On the one hand, variation in β_s across households receiving the EIP at different times could be significant because when each household received its EIP is nonrandom (unlike in previous payment programs). Later recipients tended to be households for which the IRS did not have their bank

47. In a dynamic specification where leads and lags are added, there is also the additional problem of contamination; see Sun and Abraham (2021) for details.

information or physical address and so have slightly lower incomes and expenditures on average. In addition, the pandemic period was a period of unprecedented economic volatility, and variation in β_s over time could arise from variation in the economy or the pandemic situation.[48] On the other hand, most of our variation comes from comparing recipients to non-recipients (always a valid comparison) and comparing people receiving different amounts of EIPs. Further, Parker and others (2021) show through simulation that there is minimal bias for quite substantial variation in average treatment effect over time for the first round of EIPs, where the variation in timing of receipt is the greatest of the three.

Our solution is to follow Borusyak, Jaravel, and Spiess (2021), which allows differences in MPC or β_s over time and is unbiased under generalized parallel trends (and no treatment anticipation) assumptions.[49] The estimation method can be clearly described as a three-step procedure. Denoting the set of never-treated and not-yet-treated observations as Ω_0, in the first step we estimate the time dummies and coefficients on controls using only Ω_0:[50]

$$\Delta\tilde{C}_{i,t} = \tilde{X}_{i,t}\gamma + \tau_t + \alpha_{R(i)} + \eta_{i,t} \quad \forall\{i,t\} \in \Omega_0. \tag{3}$$

In the second step, for treated observations only, we compute the difference between observed scaled change in expenditure and the scaled change in expenditure predicted by controls and time, denoted by $\Delta\check{C}_{i,t}$:

$$\Delta\check{C}_{i,t} = \Delta\tilde{C}_{i,t} - \tilde{X}_{i,t}\hat{\gamma} + \hat{\tau}_t - \hat{\alpha}_{R(i)} \quad \forall\{i,t\} \notin \Omega_0. \tag{4}$$

Thus, $\Delta\check{C}_{i,t}$ is an estimate of the household-level spending response to the EIPs. In the third step, we run a weighted least squares (WLS) regression of the new dependent variable on the EIP variable(s) of interest:

$$\Delta\check{C}_{i,t} = \sum_{s=0}^{S}\beta_s\widetilde{EIP1}_{i,t-s} + \check{\epsilon}_{i,t}. \tag{5}$$

48. Also, the CE interview structure could lead to heterogeneity. Even for households that received the payment on the same day and had the same spending response in reality, if they were interviewed in different months and hence had different reference periods, the measured spending response would differ.

49. The estimator is also efficient under homoskedasticity and is asymptotically conservative when standard errors are clustered.

50. As noted, for EIP2 analysis, $\widetilde{EIP1}_{i,t-s}$ and $\widetilde{EIP3}_{i,t-s}$ are added as controls. Similarly, for EIP3 analysis, $\widetilde{EIP1}_{i,t-s}$ and $\widetilde{EIP2}_{i,t-s}$ are added as controls.

Our method solved the issue created by "forbidden comparison," but note that the third step deviates from Borusyak, Jaravel, and Spiess (2021)—we rely on regressions to compute average MPC instead of aggregating individual effects using proposed weights. This change allows us to exploit the differences in treatment intensity and to compare different specifications. To the best of our knowledge, those features cannot yet be achieved for our specific setting by any of the new estimators to date. The disadvantage is that the weights used in the regressions are not as explicit and could be hard to interpret.[51]

To better approximate the average response, we also use the average CE weight across all interviews for each household. In practice, whether one weights or not (or whether one uses replication weights) makes very little difference to the estimates.[52]

IV. The Average MPC in Response to the Arrival of Each EIP

This section presents the results of our analysis of the spending responses to all three rounds of the EIPs using the same survey data source, the CE survey, as was used in studying the 2001 and 2008 tax payments. We show that the estimated short-term spending responses out of EIPs are small whether we use the new and improved estimation method just described or the exact same method as used in the studies of the 2001 and 2008 payments. The estimated spending responses are small both relative to the responses estimated for the past tax payments and relative to other estimates of spending responses to these EIPs that are based on other populations and data sets.

Table 3 displays the main spending responses to all three rounds of EIPs, both the average fraction of the EIP that is spent shortly after arrival

51. However, some early evidence shows that after addressing forbidden comparison, the weighting issue is unlikely to lead to significant bias since the estimate will be a convex weighted average; see, for example, Baker, Larcker, and Wang (2022) and Roth and others (2022) for the stacked regression method.

52. We make three other choices that differ slightly from previous analyses. As in previous papers, we drop the bottom 1 percent of the distribution in broad nondurable consumer expenditures after adjusting for family size, but instead of estimating the bottom 1 percent using a quantile regression on a linear trend, we drop the bottom 1 percent in each interview to account for the volatility across time during our sample due to the pandemic. Second, we do not drop households older than 85. Finally, we choose to follow panel A of table 3 in Parker and others (2013) rather than table 2, which means allowing a different average growth rate of expenditure between recipients and non-recipients. Our estimates are largely insensitive to these three choices.

Table 3. The Contemporaneous Response of Consumer Expenditures to EIP Receipt

	MPC				*Dollars spent*			
	Food and alcohol	*Strictly nondurables*	*Nondurable goods and services*	*All CE goods and services*	*Food and alcohol*	*Strictly nondurables*	*Nondurable goods and services*	*All CE goods and services*
Panel A. EIP1								
*EIP*1	0.011	0.075	0.102	0.234				
	(0.016)	(0.020)	(0.028)	(0.059)				
1[*EIP*1 > 0]					6.5	96.4	80.8	336.5
					(25.3)	(36.6)	(46.4)	(96.6)
Panel B. EIP2								
*EIP*2	0.034	0.103	0.083	0.247				
	(0.021)	(0.031)	(0.039)	(0.090)				
1[*EIP*2 > 0]					18.8	80.8	65.6	156.7
					(23.6)	(44.0)	(52.2)	(114.4)
Panel C. EIP3								
*EIP*3	0.036	0.030	0.009	0.015				
	(0.017)	(0.016)	(0.018)	(0.043)				
1[*EIP*3 > 0]					99.5	86.8	55.1	–36.0
					(33.8)	(40.8)	(42.2)	(102.4)
Average quarterly household spending across three waves								
	$2,292	$4,516	$5,996	$14,401	$2,292	$4,516	$5,996	$14,401

Source: Authors' calculations.

Note: Table reports estimation of equations (3)–(5) with $S = 1$, with scaled dollar change in consumption as the dependent variable and using weighted least squares using average weights. Each pair of rows uses the final sample for that EIP round. Standard errors included in parentheses are adjusted for arbitrary within-household correlations and heteroskedasticity. Besides separate intercepts, regressions also include interview month dummies, scaled age and change in the size of the CU, and controls for the other EIPs for EIP2 and EIP3. For EIP1, the four columns have 3,541, 3,543, 3,543, and 3,544 treated observations, and 2,264 never-treated or not-yet-treated observations except for the first column, which has 2,261 observations. For EIP2, the columns have 3,171, 3,171, 3,175, and 3,175 treated observations, and 5,002, 5,004, 5,004, and 5,005 never-treated or not-yet-treated observations. For EIP3, the columns have 3,566, 3,566, 3,568, and 3,567 treated observations, and 3,465, 3,474, 3,477, and 3,474 never-treated or not-yet-treated observations.

(first four columns) and the average dollar amount that is spent (last four columns). These results come from our main estimation method of equation (2) (the three-step, unbiased procedure) with $S = 1$.

The first four columns of the first row of panel A show that the first round of EIPs was not spent rapidly after receipt and so on average was not providing urgently needed pandemic insurance. During the three-month reference period in which a payment was received, a household on average increased its spending on nondurable goods and services by 10.2 percent of EIP1, and on all CE-measured goods and services by 23.4 percent of EIP1. Taking the perspective of classical statistics, the 95 percent confidence intervals of cumulative spending rule out spending in excess of 16 percent of the EIP on nondurable goods and services and 35 percent on all CE goods and services.

The first four rows of panel B show similar low spending responses for the second round of EIPs. The third and fourth columns show that 8 percent and 25 percent of the EIP2s were used for expenditures on nondurable goods and services and total CE-measured expenditures, respectively, within the three-month period of receipt. These first two panels are consistent with the hypothesis that, because households tilted spending toward durable goods during the pandemic, the spending response to the EIPs was similarly tilted toward durable goods. Compared to past stimulus programs, the share of spending going to durable goods does appear higher than in 2001, but it is not higher than in 2008, and the statistical strength of both comparisons is weak.

Finally, the first four rows of panel C show even lower spending responses for the third round of EIPs than for the first and second rounds. Spending in response to EIP3 receipt was economically (and statistically) close to zero. As noted, because it is possible that some households that received EIP2 or EIP3 payments failed to report them, one should maintain some skepticism that the actual spending response was quite this low, particularly for the third round of the EIPs. However, the lower spending response is consistent both with the rise in liquid balances throughout the pandemic (Grieg, Deadman, and Sonthalia 2021) and with the large dollar size of the third round of the EIPs.

How might our estimated spending response to EIP2 and EIP3 be lowered by underreporting of EIP receipt in the CE? Underreporting implies that some households in our control group were actually treated and so reduces the difference we measure between groups. To shed light on this possibility, we calculate EIP receipt and amount from the rules of each round of EIP and the TAXSIM imputations contained in the CE as described in online appendix A.6. We create alternative measures of our EIP variable for each round of EIP by assuming that any EIP arrived in the first two months of that round. We then conduct our main analysis using these imputed EIPs and dropping any CE household with a recall period that does not contain both of the critical two months. Online appendix tables C.8–C.10 show the results of our analysis. Using imputed EIPs, the estimated MPCs for EIP2 are smaller than those in our main analysis, suggesting we are not overestimating the spending response in the second round. For the third round, however, analysis of these alternative measures suggests that our estimated MPCs out of EIP3 are indeed underestimated, but this alternative analysis still finds them to be relatively small.

The last four columns of table 3 show the dollar spending response to receipt of an EIP (rather than the MPC) and imply smaller spending responses. These columns are based on our main estimation but replacing our measure of EIP amount with an indicator variable of EIP1 receipt, $\mathbb{1}[EIPn_{i,t-s} > 0]$ so that these estimates do not identify the spending effect using any information about EIP amounts across recipients. The estimated dollar spending responses to the arrival of EIP1 are \$81 or 3 percent of the average EIP1 on nondurable goods and services (statistically insignificant, column 7) and \$337 or 16 percent of the average EIP1 on all measure CE expenditures (statistically significant, column 8). For EIP2 the spending responses of \$66 and \$157, respectively (statistically insignificant), are 5 percent and 12 percent of the average EIP2 and so imply even less spending than the specification in the first four columns. Finally, the last four columns also continue to show very small spending responses to the third round of the EIPs, particularly because the average EIP3 is one-third bigger than the average EIP1.

We have measured EIP-driven spending in the short term to evaluate whether the EIPs provided urgently needed pandemic insurance, and we now turn to evaluating subsequent spending, which is informative both about pandemic insurance but over a three-month-longer period and, for longer horizons, about the contribution of EIPs to the rapid pandemic recovery and potentially inflation. In terms of pandemic insurance, we find some evidence of continued higher spending for EIP1 and EIP2 but no evidence of any continued spending for EIP3. In terms of increases in demand over any longer periods, we lack the statistical power to add any evidence on the potential contribution of EIPs to strong demand or inflation during the second half of 2021 or beyond.

Table 4 shows the longer-run response of spending to the receipt of an EIP. The coefficient β_1 on $\widetilde{EIP}_{i,t-s}$ measures the decline in spending during the three months following receipt, so that $\beta_0 + \beta_1$ measures the increase in spending in the second three months relative to the previous three months. The bottom row of the table reports $\beta_0 + (\beta_0 + \beta_1)$, the sum of the contemporaneous spending and this additional spending, which is then the total spending during both the three-month period of receipt and the subsequent three-month period (as a percentage of the EIP).

For EIP1, the cumulative MPC on strictly nondurable and broad nondurable goods and services are both roughly 13 percent and on all CE goods and services is 45 percent (with a standard error of 15.8 percent). For EIP2, the MPCs are slightly higher, consistent with the more open economy and the smaller size of the payments. Finally, for EIP3, we find no evidence that EIP3s were spent during the three months of receipt or during the subsequent three-month period. Online appendix tables C.11–C.13 show that using our imputed EIP measures described above does not change the conclusion of small spending effects.

Table 5 summarizes our finding of low spending response to these EIPs and compares the spending responses to those of earlier stimulus payment programs. The MPCs out of the EIPs are substantively lower than MPCs out of tax payments disbursed in 2001 and 2008, according to studies using the same survey data.

Are these relatively low spending responses due to our differences (improvements) in methodology? No. To show this, we apply the methodology of Johnson, Parker, and Souleles (2006) and Parker and others (2013) exactly and estimate spending responses to each round of EIPs on the sample of all CE households. The estimated spending responses are unstable across specifications and columns but on average are not inconsistent with the results shown in table 3 for EIP1 and EIP3 (results for EIP2 suggest even smaller spending responses).

More precisely, we estimate equation (1) on samples that are constructed exactly as in these earlier papers, and replicate table 2 in both of these papers, for all three rounds of EIPs. As shown in the first four columns of table 6, for EIP1 point estimates suggest MPCs of 4.3 percent on food, 7.1 percent on strictly nondurables, 7.7 percent on the broad measure of nondurable goods and services, and 28.0 percent on all goods and services. While all these estimates are statistically insignificant, these point estimates are consistent with those in table 3. But this methodology leads to wildly different conclusions for other specifications (unlike when the same analysis was used on the 2001 and 2008 tax payments), consistent with

Table 4. The Longer-Term Response of Consumer Expenditures to EIP Receipt

	Dependent variable: scaled dollar change in spending on:								
	EIP1			*EIP2*			*EIP3*		
	Strictly nondurables	*Nondurables*	*All CE goods and services*	*Strictly nondurables*	*Nondurables*	*All CE goods and services*	*Strictly nondurables*	*Nondurables*	*All CE goods and services*
$EIPn_t$	0.075	0.102	0.234	0.103	0.083	0.247	0.030	0.009	0.015
	(0.020)	(0.028)	(0.059)	(0.031)	(0.039)	(0.090)	(0.016)	(0.018)	(0.043)
$EIPn_{t-1}$	−0.011	−0.080	−0.017	0.030	−0.013	0.107	0.000	−0.049	−0.150
	(0.020)	(0.028)	(0.070)	(0.038)	(0.045)	(0.124)	(0.010)	(0.019)	(0.049)
Implied cumulative fraction of EIP spent over two three-month periods									
	0.139	0.124	0.452	0.235	0.153	0.601	0.059	−0.030	−0.119
	(0.051)	(0.068)	(0.158)	(0.086)	(0.104)	(0.257)	(0.036)	(0.047)	(0.112)

Source: Authors' calculations.

Note: Table reports β_0 and β_1 from estimation of equations (3)–(5) with $S = 1$. Regressions also include interview month dummies, a separate intercept for non-recipients, scaled age, and change in the size of the CU. Panels B and C additionally control for the other EIP waves. The sample is the final sample which includes only CE households with income that does not exceed a certain threshold determined by marital status and family structure. Regressions are conducted using weighted least squares, where the weights applied are average weights. Standard errors included in parentheses are adjusted for arbitrary within-household correlations and heteroskedasticity. For EIP1, observations are those with an interview in June or July 2020; the columns have 2,264 never-treated or not-yet-treated observations and 3,543 treated observations. For EIP2, observations are those with an interview in February, March, or April 2021; the columns have 4,815, 4,817, and 4,818 never-treated or not-yet-treated observations and 3,171, 3,175, and 3,175 treated observations, respectively. For EIP3, observations are those with an interview in April, May, or June 2021; the columns have 3,474, 3,477, and 3,474 never-treated or not-yet-treated observations, and 3,566, 3,568, and 3,568 treated observations, respectively.

Table 5. Estimated MPCs on CE-Measured Nondurable Goods and Some Services

	Full sample, three months of receipt	*Recipients only, three months of receipt*	*Full sample, three months of receipt and subsequent three months*
2001 economic rebates	0.386	0.247	0.691*
	(0.135)	(0.213)	(0.260)
2008 stimulus payments	0.121	0.308	0.347
	(0.055)	(0.112)	(0.155)
2020 EIP1	0.102	−0.062	0.124
	(0.028)	(0.072)	(0.068)
2020 EIP2	0.083		0.153
	(0.039)		(0.104)
2020 EIP3	0.009		−0.030
	(0.018)		(0.047)

Sources: Johnson, Parker, and Souleles (2006); Parker and others (2013); and Parker and others (2021).

Note: The asterisk (*) denotes a large MPC driven in part by one outlier in spending on food.

the arguments for our preferred specification in section III. The last four columns of panel A show estimates using an indicator variable for receipt in place of EIP1 amount and implies that, in the three months in which the EIP1 arrives, spending increases by \$157 on food, \$296 on strictly nondurables, \$375 on nondurables, and \$1,279 on all goods and services, with all but the first being statistically significant. For the average EIP1, these estimates would imply MPCs of 7 percent, 14 percent, 18 percent, and 61 percent, respectively, roughly double those from estimating the MPC directly (the average of $EIP_{i,t}$ conditional on receipt is \$2,098). Online appendix table C.4 shows the results of estimation for the two other specifications used in previous research, and these estimated spending responses are all statistically insignificant and again imply quite different MPCs than table 6.[53]

53. Johnson, Parker, and Souleles (2006) and Parker and others (2013) both report estimates of MPCs (in table 3) that rely only on variation in time of receipt by dropping all households that never receive stimulus payments. In these earlier episodes this variation was closer to purely random. Given the lack of variation in timing in the EIP programs, estimates of the MPC in analogous samples that drop households that never receive EIPs have very large standard errors. For EIP1, the program with the largest variation in timing of disbursement, appendix table C.1 in Parker and others (2021) shows that the standard errors are typically 50 percent to 100 percent larger than in table 6 here and online appendix C.4, as expected given the lack of variation. Additionally, the estimates are more variable and many are negative; so, we learn little from this exercise.

Table 6. **The Response of Consumer Expenditure to EIP Arrival Estimated on Recipients and Non-recipients Using the Methodology Previously Applied to Tax Rebates**

	MPC				*Dollars spent*			
	Food and alcohol	*Strictly nondurables*	*Nondurable goods and services*	*All CE goods and services*	*Food and alcohol*	*Strictly nondurables*	*Nondurable goods and services*	*All CE goods and services*
Panel A. EIP1								
$EIP1$	0.043	0.071	0.077	0.280				
	(0.032)	(0.044)	(0.059)	(0.217)				
$1[EIP1 > 0]$					157.3	296.4	375.0	1278.8
					(89.9)	(130.2)	(167.8)	(647.5)
Panel B. EIP2								
$EIP2$	0.011	0.037	0.030	0.008				
	(0.029)	(0.044)	(0.055)	(0.325)				
$1[EIP2 > 0]$					−57.1	−11.1	−10.1	−498.7
					(51.7)	(79.3)	(99.5)	(749.8)
Panel C. EIP3								
$EIP3$	0.001	0.001	0.005	0.222				
	(0.013)	(0.017)	(0.023)	(0.149)				
$1[EIP3 > 0]$					14.2	−6.3	22.7	702.1
					(45.1)	(70.3)	(91.4)	(648.7)

Source: Authors' calculations.

Note: Table reports β_0 from estimation of equation (1) with $S = 0$ with dollar change in consumption as the dependent variable and weighted least squares using average weights. Standard errors included in parentheses are adjusted for arbitrary within-household correlations and heteroskedasticity. Regressions also include interview month dummies, age, and change in the size of the CU. The samples are constructed as in previous research papers (see online appendix). Panel A has 5,634 observations and includes the sample of all CE households with an interview in June or July 2020. Panel B has 8,302 observations, includes the sample of all CE households with an interview in February, March, or April 2021, and additionally includes controls for EIP1 and EIP3. Panel C has 7,335 observations, includes the sample of all CE households with an interview in April, May, or June 2021, and additionally includes controls for EIP1 and EIP2.

V. EIPs as Pandemic Insurance

While we find low average spending responses to the EIPs, the EIPs may nonetheless have filled urgent economic needs for some subset of households, presumably those who experienced the greatest impact economically as a result of the pandemic. In this section, we construct observable measures of economic vulnerability to the economic consequences of the pandemic and evaluate whether households that were more exposed spent more of their EIPs to maintain or increase their consumer spending in the short run. We focus both on households with little ex ante liquid wealth and on households with labor income exposed to the pandemic as measured from their ability to work from home. While the average spending response to the EIPs is low, consistent with payments not being required to fill short-term spending needs for most households, we find two pieces of evidence that the EIPs did raise spending and so provided potentially important assistance to some households. First, we show that households entering the pandemic period with little ex ante liquid wealth spent a larger share of their EIP1s. For EIP2 and EIP3, there is little to no evidence that households with low liquid wealth had higher MPCs. Second, we show that households whose incomes were more exposed to the pandemic—those with lower ability to work from home—spent more out of their first-round EIPs when they arrived. For the second round of EIPs we find no such pattern of MPC related to the ability to work from home. For the third round, there is some evidence of a small effect.

We estimate different MPCs for different groups of recipients by interacting the EIP variables in equation (2) with a group membership indicator variable, denoted $g(i)$, so that the spending response of interest varies by group as well as horizon. We use the equation:

$$\Delta \tilde{C}_{i,t} = \sum_{s=0}^{S} \beta_{g(i),s} \widetilde{EIPn}_{i,t-s} + \tilde{X}_{i,t}\gamma + \alpha_{g(i)} + \tau_t + \epsilon_{i,t}, \tag{6}$$

which also allows the intercept or average growth rate of spending to differ by group ($\alpha_{g(i)}$). For studying the MPC of EIP2 and EIP3, we also interact the controls for other EIPs (in X) with the indicator for group membership. To be clear, consider the MPC for EIP2. We estimate equation (6) using our imputation estimator and the procedure described in equations (3)–(5).

First, we split the sample of households by their ex ante liquid wealth and find that households that entered the pandemic with low liquidity had strong spending responses to the first round of EIPs in the CARES Act.

We measure liquid wealth as the sum of balances in checking accounts, saving accounts, money market accounts, and certificates of deposits at the start of the households' first interview (reported in the last interview).[54] Table 7 shows that, for EIP1, households in the bottom third of the distribution of liquidity—those with less than $2,000 available, which is still a substantial amount—have statistically significant MPCs of 6 percent, 22 percent, and 48 percent on food, nondurable goods and services, and all CE goods and services, respectively. While the difference between each of these MPCs and the corresponding MPC of either of the other third of the distribution is not statistically significant, they are economically large, and we can reject the equality of MPCs across these three groups for spending on both nondurable goods and services and all CE goods and services.

Previous research on tax rebates that uses the CE survey has not consistently found a statistically significant decreasing relationship between spending responses and liquidity. However, analyses with better measures of liquidity have generally found a larger MPC for households with lower liquidity (Parker 2017; Olafsson and Pagel 2018; Ganong and others 2020; Baugh and others 2021; Fagereng, Holm, and Natvik 2021).

For the second round of EIPs, the spending responses are higher for households in the bottom two thirds of the liquidity distribution, and we can no longer reject equality of the MPCs across the thirds of the distribution of liquid wealth. No spending responses are statistically significant, but point estimates suggest the least liquid households spent 12 percent of their EIPs on nondurable goods and services, the middle third in terms of liquidity spent 11 percent, while the most liquid households are estimated to spend a negative amount. The MPCs on total expenditures are more related to liquidity: 41 percent, 22 percent, and −5 percent as we move from the lowest to highest third of the distribution of liquid wealth but again with no estimate being statistically significant. These findings are not inconsistent with Garner and Schild (2021), which shows that in the Household

54. Even the low liquidity group has substantial reported wealth, and in particular the distribution of reported liquid wealth is much higher in these 2020 data than it was in 2008. In Parker and others (2013) the 33rd percentile in the distribution of liquid wealth was only $500. One possibility is changes in the distribution of respondents, although this appears unlikely, as we discuss in online appendix A.4. More likely, this difference reflects changes in the CE survey and the financial accounts that it covers. In 2008 the CE asked about balances in checking and saving accounts separately, but in 2013 the CE survey switched to asking a single question about total liquidity across a larger set of types of accounts, and starting in 2017 the survey introduced an initial question asking whether there was a zero balance in these accounts. The latter change was associated with a reduction in the number of households reporting zero balances.

Table 7. The Contemporaneous Response of Consumer Expenditures to EIP by Liquidity

	Dependent variable: scaled dollar change in spending on:								
	EIP1			*EIP2*			*EIP3*		
	Bottom third ≤ *2,000* *Top third* ≥ *12,667*			*Bottom third* ≤ *2,000* *Top third* ≥ *12,000*			*Bottom third* ≤ *2,000* *Top third* ≥ *10,000*		
	Food and alcohol	*Nondurables*	*All CE goods and services*	*Food and alcohol*	*Nondurables*	*All CE goods and services*	*Food and alcohol*	*Nondurables*	*All CE goods and services*
$EIPn_t$	0.039	0.087	0.178	−0.032	0.112	0.220	0.078	0.132	0.081
	(0.033)	(0.064)	(0.155)	(0.071)	(0.111)	(0.326)	(0.035)	(0.062)	(0.197)
$EIPn_t$ × *Bottom third*	0.016	0.130	0.301	0.050	0.009	0.191	−0.018	−0.099	−0.065
	(0.051)	(0.095)	(0.210)	(0.084)	(0.157)	(0.403)	(0.048)	(0.087)	(0.219)
$EIPn_t$ × *Top third*	0.013	−0.188	−0.275	−0.090	−0.255	−0.272	0.048	−0.129	−0.057
	(0.046)	(0.102)	(0.243)	(0.085)	(0.139)	(0.449)	(0.101)	(0.099)	(0.267)
p-value for test of equality of responses	0.942	0.011	0.044	0.107	0.077	0.492	0.784	0.355	0.957
Implied propensity to spend by group									
Least liquid third	0.055	0.217	0.479	0.018	0.121	0.411	0.060	0.033	0.016
	(0.039)	(0.070)	(0.142)	(0.046)	(0.112)	(0.237)	(0.034)	(0.062)	(0.095)
Most liquid third	0.052	−0.101	−0.097	−0.122	−0.143	−0.052	0.126	0.003	0.024
	(0.032)	(0.079)	(0.188)	(0.048)	(0.083)	(0.309)	(0.094)	(0.077)	(0.180)

Source: Authors' calculations.

Note: All regressions use our imputation estimator to estimate equation (6). Also included are interview month dummies, scaled age and change in the size of the CU, and separate intercepts by thirds of the liquidity distribution interacted with other EIPs. The sample is the final sample, and all results are from WLS regressions using average weights. Standard errors included in parentheses are adjusted for arbitrary within-household correlations and heteroskedasticity. The tests of equal responses are joint test for H_0:$\beta_{0,Bottom\ third} = 0$ and $\beta_{0,Top\ third} = 0$. For EIP1, observations are those with an interview in June or July 2020; the columns have 1,065, 1,066, and 1,066 never-treated or not-yet-treated observations, and 1,608, 1,609, and 1,609 treated observations, respectively. For EIP2, observations are those with an interview in February, March, or April 2021; the columns have 1,795 never-treated or not-yet-treated observations and 1,211 treated observations. For EIP3, observations are those with an interview in April, May, or June 2021; the columns have 1,387, 1,390, and 1,387 never-treated or not-yet-treated observations, and 892 treated observations.

Pulse data, households reporting higher levels of financial difficulty are more likely to use their EIP2s mostly for spending.

Finally, for the third round of EIP—the largest in dollar terms, the latest in the pandemic, and the most likely to be understated due to data issues—the middle of the liquidity distribution is the only group estimated to have a statistically significant spending response to the arrival of their EIPs: 13 percent (6 percent) on nondurable goods and services, compared to 3 percent (6 percent) and 0.3 percent (8 percent) for the bottom and top thirds of the distribution of liquid wealth, respectively. Again, we cannot reject the null hypothesis of no differential response.

These patterns suggest that the first round of EIPs did meet important liquidity needs for households with little liquid wealth in the early stages of the pandemic, when the economy was most shut down. But later EIP rounds appear less beneficial on this front (or their benefits were less related to liquid wealth). The second-round payments were broadly spent at the same average rate as EIP1, consistent with the tendency for households to spend out of small, transitory increases in liquidity, and also with similar constraints on consumer spending from the pandemic as EIP1. And the low spending of the final round of payments, particularly among households with little liquidity is consistent with the large size of the payment, although again our caveat about the low rate of EIP receipt reported in the CE survey applies.

Analysis of our second measure of whether the EIPs provided effective pandemic insurance—based on households' ability to work from home—paints a similar picture: the first round of EIPs appears to fill a pandemic insurance need for households but later rounds do not.

We measure the exposure of income to the inability to work from home for EIP1 by the share of pre-pandemic household income that cannot be earned from home. Specifically, for the reference person and any secondary earner, we calculate the share of tasks associated with their job based on their industry and education level following a mapping into the classifications by occupation and education in Mongey, Pilossoph, and Weinberg (2021) and Dingel and Neiman (2020). For individuals with no earned income (valid missing earnings), like retirees or people not in the labor force, the measure is zero. We then multiply this share by each person's wage and salary income, sum to the household level, and divide by family income. Because we require pre-pandemic income, we only use this measure to analyze EIP1. Online appendix B.3 contains complete details.

Table 8 shows that households most reliant on labor income from jobs that cannot be done at home account for most of the spending response

Table 8. The Response of Consumer Expenditures to EIP1 Receipt by the Exposure of Income to Inability to Work from Home in 2020

	Dependent variable: scaled dollar change in spending on:		
	Food and alcohol	*Nondurables*	*All CE goods and services*
Fraction of EIP1 spent over contemporaneous three-month period			
$EIP1_t$	0.021	0.052	−0.049
	(0.022)	(0.055)	(0.119)
$EIP1_t \times$ *Middle third*	0.030	0.176	0.258
	(0.043)	(0.089)	(0.232)
$EIP1_t \times$ *Least able third*	0.036	0.064	0.367
	(0.038)	(0.083)	(0.188)
p-value for test of equality of responses	0.731	0.225	0.210
Cumulative fraction of EIP1 spent over contemporaneous and next three-month period			
Most able third	−0.007	−0.135	−0.435
	(0.057)	(0.159)	(0.349)
Middle third	0.126	0.365	0.181
	(0.100)	(0.190)	(0.622)
Least able third	0.117	0.285	0.842
	(0.080)	(0.156)	(0.448)

Source: Authors' calculations.

Note: All regressions use equation (6). Also included are interview month dummies, scaled age and change in the size of the CU, and separate intercepts by thirds of the distribution. The sample is the final sample which includes only CE households with an interview in June or July 2020, and with income that does not exceed a certain threshold determined by marital status and family structure. The work-from-home measure used is the income-based measure. All results are from WLS regressions. Weights applied are average weights. Standard errors included in parentheses are adjusted for arbitrary within-household correlations and heteroskedasticity. The tests of equal responses are joint test for H_0: $\beta_{0,Least\ able\ third} = 0$ and $\beta_{0,Middle\ third} = 0$.

to the first round of EIPs. The third of households with little to no income exposure have point estimates that imply EIP1 lowered their spending. The third of households with income that was moderately exposed had an average MPC of 37 percent (19 percent) on nondurable goods and services, while the most exposed third had a similar average MPC of 29 percent (16 percent), and an MPC on total expenditures of 84 percent (45 percent), during the three-month period of receipt and the subsequent period.

For later EIPs, given the rotating panel structure of the CE, we cannot measure pre-pandemic incomes, and earnings after the onset of the pandemic may already reflect losses incurred by an inability to work from home. Therefore, in order to investigate differences in consumption responses across ability to work from home for EIP2 and EIP3, we construct a work-from-home measure that does not rely on observing pre-pandemic wage and salary earnings. We construct a second measure based on the share of wage and salary (potential) earnings that cannot be done from home and the assumption that earners within a family have equal earnings. This measure requires only information on the industry and education of (potential) earners, whether currently working or not (see online appendix B.3 for details).

Using this second measure, table 9 shows findings for EIP1 that align well with our first measure of the ability to work from home based on pre-pandemic income. That is, we find all spending is done by the two-thirds of households with the highest level of income exposure during the pandemic, as we did in table 8. There are no significant differences in spending propensities related to the ability to work from home for either of the second two rounds of EIPs, consistent with the waning of the economic impact of the pandemic. If anything, EIP2 spending responses are concentrated among households with no income exposure to the pandemic. For EIP3, only those with incomes that are the most exposed to the pandemic have statistically significant spending response on nondurable goods and services.

In sum, while on average the EIPs appear to have gone to many households with incomes that were unharmed by the pandemic (e.g., retirees, those employed and able to work from home, etc.), some of the EIPs, mainly in the first round, did support short-term spending for some households, those with low ex ante liquid wealth and those reliant on income that could not be earned by working from home.

VI. Concluding Discussion

The pandemic limited the types of goods and services that people could purchase and many households reduced spending. There were also policy responses besides the EIPs, including extended and expanded unemployment insurance and the Paycheck Protection Program, which transferred money to small and medium-sized businesses with some incentives to maintain payroll, both of which were intended to help offset any lost income. Finally, the depth and duration of the pandemic were uncertain, particularly when the first round of EIPs was being disbursed. These factors appear to have led to less spending on nondurable goods and services (CE-measured) in response to the arrival of the first round of EIPs than was the case with the tax rebates in 2001 and 2008 and to have tilted what spending response there was toward durable goods. We observe low and similar spending responses to the first and second rounds of EIPs but very little short-run spending in response to the third round, consistent with preexisting high levels of financial resources, although the response is not as cleanly measured as the first two rounds of EIPs.

Table 9. The Contemporaneous Response of Consumer Expenditures to EIP Receipt by Ability to Work from Home

	Dependent variable: scaled dollar change in spending on:								
	EIP1			*EIP2*			*EIP3*		
	Bottom third ≤ 89.4% *Top third ≥ 99.1%*			*Bottom third ≤ 87.0%* *Top third ≥ 98.8%*			*Bottom third ≤ 84,1%* *Top third ≥ 97.7%*		
	Food and alcohol	*Nondurables*	*All CE goods and services*	*Food and alcohol*	*Nondurables*	*All CE goods and services*	*Food and alcohol*	*Nondurables*	*All CE goods and services*
$EIPn_t$	−0.014	0.048	0.189	0.061	0.131	0.265	0.026	−0.021	0.015
	(0.027)	(0.041)	(0.086)	(0.035)	(0.071)	(0.144)	(0.036)	(0.038)	(0.080)
$EIPn_t \times$ *Middle third*	0.031	0.090	0.158	−0.042	−0.131	−0.211	0.024	0.033	0.219
	(0.038)	(0.062)	(0.149)	(0.049)	(0.095)	(0.230)	(0.039)	(0.050)	(0.122)
$EIPn_t \times$ *Least able third*	0.090	0.109	0.239	−0.004	−0.024	−0.072	0.016	0.066	0.143
	(0.037)	(0.069)	(0.144)	(0.053)	(0.100)	(0.216)	(0.039)	(0.048)	(0.106)
p-value for test of equality of responses	0.046	0.188	0.222	0.648	0.326	0.656	0.817	0.376	0.174
Implied propensity to spend by group									
Middle third	0.016	0.138	0.346	0.019	0.000	0.054	0.049	0.012	0.233
	(0.026)	(0.047)	(0.121)	(0.034)	(0.063)	(0.179)	(0.015)	(0.033)	(0.093)
Least able third	0.076	0.157	0.428	0.057	0.110	0.193	0.042	0.045	0.158
	(0.026)	(0.055)	(0.116)	(0.040)	(0.070)	(0.160)	(0.016)	(0.029)	(0.070)

Source: Authors' calculations.

Note: All regressions use equation (6). Also included are interview month dummies, scaled age and change in the size of the CU, and separate intercepts by thirds of the distribution. The sample is the final sample which includes only CE households with an interview in June or July 2020, with income that does not exceed a certain threshold determined by marital status and family structure. The work-from-home measure used is the non-income measure. All results are from WLS regressions, and the weights applied are average weights. Standard errors included in parentheses are adjusted for arbitrary within-household correlations and heteroskedasticity. The tests of equal responses are joint test for H_0: $\beta_{0,\,Bottom\ third} = 0$ and $\beta_{0,Middle\ third} = 0$. For EIP1, all regressions have 3,470 observations. For EIP2, all regressions have 3,099 observations. For EIP3, all regressions have 3,463 observations.

Were the EIPs effective? The goal of previous tax rebate programs was to increase demand, and so their efficacy is largely related to the speed and size of the spending responses. In contrast, the policy goal of the EIPs was insurance, that is, to provide money to those who lost or would lose employment and who would not be covered by government aid programs. For these individuals, the EIPs could be initially saved and then used to cover a later loss. We find significant spending responses for households with low levels of ex ante liquidity in response to the first round of EIPs during the national emergency at the onset of the pandemic. The smaller amount of spending following the arrival of the December 2020 payments was due to a spending response by those outside the top third of the liquidity distribution. Finally, we find substantially higher spending responses by those reliant on earnings from jobs with tasks that could not be done from home in response to the first-round EIPs (and little evidence on this issue for later EIP rounds).

The small, short-term spending response and its pattern suggest that the EIPs went to many people who did not need the additional funds as urgent pandemic insurance.[55] However, despite the lack of much immediate spending, the EIPs could have filled the role of pandemic insurance for some households beyond the time horizon accurately measured by this (and other) studies. On the other hand, from a demand management perspective, the unspent EIPs have contributed to strong household balance sheets over the past year, a period of strong demand and rising inflation.

ACKNOWLEDGMENTS Jianmeng Lyu and Ian Sapollnik provided excellent research assistance. For useful discussions, we thank Peter Ganong, Thesia Garner, Jianmeng Lyu, Claudia Sahm, Gianluca Violante, and, in particular, Karen Dynan, Jan Eberly, and Matt Rognlie, as well as participants at the 2022 Society of Government Economists Annual Conference, the Conference on Direct Stimulus Payments to Individuals in the COVID-19 Pandemic at Athens University of Economics and Business and University of Exeter Business School (online), a Bundesbank "Friendlyfaces" workshop, and the Fall 2022 *BPEA* Conference. The initial phases of this work occurred while David S. Johnson was at the University of Michigan.

55. Sahm (2021) debates these issues.

References

Agarwal, Sumit, Chunlin Liu, and Nicholas S. Souleles. 2007. "The Reaction of Consumer Spending and Debt to Tax Rebates: Evidence from Consumer Credit Data." *Journal of Political Economy* 115, no. 6: 986–1019.

Agarwal, Sumit, and Wenlan Qian. 2014. "Consumption and Debt Response to Unanticipated Income Shocks: Evidence from a Natural Experiment in Singapore." *American Economic Review* 104, no. 12: 4205–30.

Armantier, Olivier, Leo Goldman, Gizem Koşar, Jessica Lu, Rachel Pomerantz, and Wilbert Van Der Klaauw. 2020. "How Have Households Used Their Stimulus Payments and How Would They Spend the Next?" Blog post, October 13, Liberty Street Economics, Federal Reserve Bank of New York.

Baker, Andrew C., David F. Larcker, and Charles C. Y. Wang. 2022. "How Much Should We Trust Staggered Difference-in-Differences Estimates?" *Journal of Financial Economics* 144, no. 2: 370–95.

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

Baugh, Brian, Itzhak Ben-David, Hoonsuk Park, and Jonathan A. Parker. 2021. "Asymmetric Consumption Smoothing." *American Economic Review* 111, no. 1: 192–230.

Bodkin, Ronald. 1959. "Windfall Income and Consumption." *American Economic Review* 49, no. 4: 602–14.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. "Revisiting Event Study Designs: Robust and Efficient Estimation." Arxiv:2108.12419. Ithaca, N.Y.: Cornell University.

Boutros, Michael. 2021. "Evaluating the Impact of Economic Impact Payments." Working Paper. Social Science Research Network, January 24. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3742448.

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

Cajner, Tomaz, Leland D. Crane, Ryan A. Decker, John Grigsby, Adrian Hamins-Puertolas, Erik Hurst, Christopher Kurz, and Ahu Yildirmaz. 2020. "The U.S. Labor Market during the Beginning of the Pandemic Recession." Working Paper 27159. Cambridge, Mass.: National Bureau of Economic Research.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225, no. 2: 200–230.

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

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber. 2020. “How Did US Consumers Use Their Stimulus Payments?” Working Paper 27693. Cambridge, Mass.: National Bureau of Economic Research.

Cooper, Daniel H., and Giovanni P. Olivei. 2021. “High-Frequency Spending Responses to Government Transfer Payments.” Working Paper 21-10. Federal Reserve Bank of Boston. https://doi.org/10.29412/res.wp.2021.10.

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

de Chaisemartin, Clément, and Xavier D’Haultfuille. 2020. “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects.” *American Economic Review* 110, no. 9: 2964–96.

Dingel, Jonathan I., and Brent Neiman. 2020. “How Many Jobs Can Be Done at Home?” *Journal of Public Economics* 189:104235.

Fagereng, Andreas, Martin B. Holm, and Gisle J. Natvik. 2021. “MPC Heterogeneity and Household Balance Sheets.” *American Economic Journal: Macroeconomics* 13, no. 4: 1–54.

Farrell, Diana, Fiona Greig, and Amar Hamoudi. 2019. *Tax Time: How Families Manage Tax Refunds and Payments*. Technical Report. New York: JPMorgan Chase Institute.

Feldman, Naomi, and Ori Heffetz. 2022. “A Grant to Every Citizen: Survey Evidence of the Impact of a Direct Government Payment in Israel.” *National Tax Journal* 75, no. 2: 229–63.

Ganong, Peter, Fiona E. Greig, Pascal J. Noel, Daniel M. Sullivan, and Joseph S. Vavra. 2022. “Spending and Job-Finding Impacts of Expanded Unemployment Benefits: Evidence from Administrative Micro Data.” Working Paper 30315. Cambridge, Mass.: National Bureau of Economic Research.

Ganong, Peter, Damon Jones, Pascal J. Noel, Fiona E. Greig, Diana Farrell, and Chris Wheat. 2020. “Wealth, Race, and Consumption Smoothing of Typical Income Shocks.” Working Paper 27552. Cambridge, Mass.: National Bureau of Economic Research.

Garner, Thesia I., Adam Safir, and Jake Schild. 2020. “Receipt and Use of Stimulus Payments in the Time of the Covid-19 Pandemic.” *Beyond the Numbers* 9, no. 10. https://www.bls.gov/opub/btn/volume-9/receipt-and-use-of-stimulus-payments-in-the-time-of-the-covid-19-pandemic.htm.

Garner, Thesia I., and Jake Schild. 2021. “Consumer Response to Economic Impact Payments during the Covid-19 Pandemic and the Role of Subjective Assessments of Well-Being: A View from the US Using a Rapid Response Survey.” Working Paper 540. Washington: Bureau of Labor Statistics. https://www.bls.gov/osmr/research-papers/2021/ec210060.htm.

Gelatt, Julia, Randy Capps, and Michael Fix. 2021. “Nearly 3 Million U.S. Citizens and Legal Immigrants Initially Excluded under the CARES Act Are Covered under the December 2020 COVID-19 Stimulus.” Washington: Migration Policy Institute.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225, no. 2: 254–77.

Greenwood, Robin, Toomas Laarits, and Jeffrey Wurgler. 2022. "Stock Market Stimulus." Working Paper 29827. Cambridge, Mass.: National Bureau of Economic Research.

Greig, Fiona, Erica Deadman, and Tanya Sonthalia. 2021. "Household Finances Pulse: Cash Balances during COVID-19." New York: JP Morgan Chase.

Guerrieri, Veronica, Guido Lorenzoni, Ludwig Straub, and Iván Werning. 2022. "Macroeconomic Implications of COVID-19: Can Negative Supply Shocks Cause Demand Shortages?" *American Economic Review* 112, no. 5: 1437–74.

Internal Revenue Service (IRS). 2020. "Treasury, IRS Release Latest State-by-State Economic Impact Payment Figures for May 22, 2020." News Release, May 22. https://www.irs.gov/newsroom/treasury-irs-release-latest-state-by-state-economic-impact-payment-figures-for-may-22-2020.

Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles. 2006. "Household Expenditure and the Income Tax Rebates of 2001." *American Economic Review* 96, no. 5: 1589–610.

Karger, Ezra, and Aastha Rajan. 2021. "Heterogeneity in the Marginal Propensity to Consume: Evidence from Covid-19 Stimulus Payments." Working Paper 2020-15. Federal Reserve Bank of Chicago.

Kim, Seonghoon, Kanghyock Koh, and Wonjun Lyou. 2020. "Do COVID-19 Stimulus Payments Stimulate the Economy? Evidence from Card Transaction Data in South Korea." Working Paper. Social Science Research Network, September 29. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3701676.

Kubota, So, Koichiro Onishi, and Yuta Toyama. 2021. "Consumption Responses to COVID-19 Payments: Evidence from a Natural Experiment and Bank Account Data." *Journal of Economic Behavior and Organization* 188:1–17.

Kueng, Lorenz. 2018. "Excess Sensitivity of High-Income Consumers." *Quarterly Journal of Economics* 133, no. 4: 1693–751.

Lewis, Daniel J., Davide Melcangi, and Laura Pilossoph. 2019. "Latent Heterogeneity in the Marginal Propensity to Consume." Staff Report 902. Federal Reserve Bank of New York.

Lusardi, Annamaria. 1996. "Permanent Income, Current Income, and Consumption: Evidence from Two Panel Data Sets." *Journal of Business and Economic Statistics* 14, no. 1: 81–90.

Meyer, Michelle, and Anna Zhou. 2020. "COVID-19 and the Consumer: Data through April 16." Data sheet, April 22. Bank of America Global Research Data Analytics.

Misra, Kanishka, Vishal Singh, and Qianyun Poppy Zhang. 2021. "Impact of Stay-at-Home-Orders and Cost-of-Living on Stimulus Response: Evidence from the CARES Act." Working Paper. Social Science Research Network, August 6. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3663493.

Mongey, Simon, Laura Pilossoph, and Alexander Weinberg. 2021. "Which Workers Bear the Burden of Social Distancing?" *Journal of Economic Inequality* 19:509–26.

Murphy, Dan. 2021. *Economic Impact Payments: Uses, Payment Methods, and Costs to Recipients*. Washington: Economic Studies at Brookings, Brookings Institution. https://www.brookings.edu/research/economic-impact-payments-uses-payment-methods-and-costs-to-recipients/.

Olafsson, Arna, and Michaela Pagel. 2018. "The Liquid Hand-to-Mouth: Evidence from Personal Finance Management Software." *Review of Financial Studies* 31, no. 11: 4398–446.

Orchard, Jacob, Valerie A. Ramey, and Johannes F. Wieland. 2022. "Micro MPCs and Macro Counterfactuals: The Case of the 2008 Rebates." Working Paper. https://econweb.ucsd.edu/~vramey/research/Micro_MPCs_and_Macro_Counterfactuals.pdf.

Parker, Jonathan A. 1999. "The Reaction of Household Consumption to Predictable Changes in Social Security Taxes." *American Economic Review* 89, no. 4: 959–73.

Parker, Jonathan A. 2017. "Why Don't Households Smooth Consumption? Evidence from a $25 Million Experiment." *American Economic Journal: Macroeconomics* 9, no. 4: 153–83.

Parker, Jonathan A., Jake Schild, Laura Erhard, and David Johnson. 2021. "Household Spending Responses to the Economic Impact of 2020: Evidence from the Consumer Expenditure Survey." Working Paper 544. Washington: US Bureau of Labor Statistics.

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

Romer, Christina D., and David H. Romer. 2022. "A Social Insurance Perspective on Pandemic Fiscal Policy: Implications for Unemployment Insurance and Hazard Pay." *Journal of Economic Perspectives* 36, no. 2: 3–28.

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. 2022. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." Arxiv:2201.01194. Ithaca, N.Y.: Cornell University.

Sahm, Claudia R. 2021. "Robust Evidence for $1400 Checks: Relief for Families and Recovery for All." Policy Brief. New York: Jain Family Institute.

Sahm, Claudia R., Matthew D. Shapiro, and Joel Slemrod. 2010. "Household Response to the 2008 Tax Rebate: Survey Evidence and Aggregate Implications." *Tax Policy and the Economy* 24, no. 1: 69–110.

Sahm, Claudia R., Matthew D. Shapiro, and Joel Slemrod. 2020. "Consumer Response to the Coronavirus Stimulus Programs." Slides. http://macromomblog.com/wp-content/uploads/2021/02/Rebates-2020-111020.pdf.

Shapiro, Matthew D., and Joel Slemrod. 1995. "Consumer Response to the Timing of Income: Evidence from a Change in Tax Withholding." *American Economic Review* 85, no. 1: 274–83.

Shapiro, Matthew D., and Joel Slemrod. 2009. "Did the 2008 Tax Rebates Stimulate Spending?" *American Economic Review* 99, no. 2: 374–79.

Souleles, Nicholas S. 1999. "The Response of Household Consumption to Income Tax Refunds." *American Economic Review* 89, no. 4: 947–58.

Splinter, David. 2022. "Stimulus Checks: True-Up and Safe Harbor Costs." Working Paper. http://www.davidsplinter.com/splinter-stimuluschecks.pdf.

Stephens, Melvin, Jr. 2003. "'3rd of tha Month': Do Social Security Recipients Smooth Consumption between Checks?" *American Economic Review* 93, no. 1: 406–22.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* 225, no. 2: 175–99.

Wooldridge, Jeffrey M. 2021. "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators." Working Paper. Social Science Research Network, August 18. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3906345.

Comments and Discussion

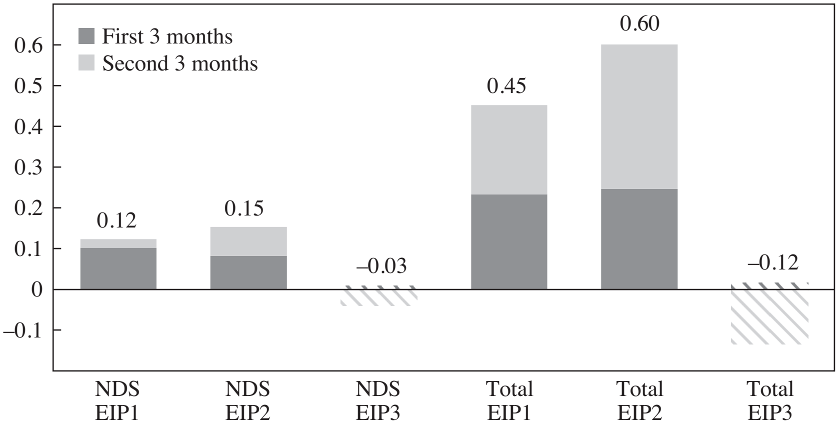
COMMENT BY

KAREN DYNAN The United States mounted a massive fiscal response to the onset of the COVID-19 pandemic in March 2020, with a key goal being to limit the economic fallout from an unprecedented shutdown of the economy. More fiscal support was passed over the subsequent year as the pandemic persisted and continued to disrupt economic activity. An important piece of this response was the three waves of direct payments to households, or Economic Impact Payments (EIPs), legislated in March 2020, December 2020, and March 2021. Economic research on how these payments affected households is crucial to designing policies that will be effective at fighting future recessions. Parker, Schild, Erhard, and Johnson provide a thoughtful analysis that contributes to an important emerging literature on this question.

This paper, like nearly all studies of the EIPs to date, explores the response of household spending to the EIPs. It stands out for its use of data from a survey of households, the Consumer Expenditure (CE) Survey, as opposed to the administrative financial records that have been the basis of most of the other studies. The paper thus represents an important complement to the rest of the literature. One advantage of the CE survey is that respondents are asked to report all types of expenditures, allowing for the construction of a very comprehensive measure of consumption, whereas studies based on credit and debit card records, for example, cannot tell you about types of spending for which cards are not typically used, such as motor vehicle purchases. The CE survey also has a more representative sample than many sources of financial records, and being able to look at households from all

Brookings Papers on Economic Activity, Fall 2022: 131–156 © 2023 The Brookings Institution.

Figure 1. Point Estimates of Cumulative Marginal Propensity to Consume



Source: Author's calculations, using data from table 4 on p. 115.
Note: Bars labeled "NDS" show estimates of the response of nondurables and services consumption to the first, second, and third rounds of EIPs; bars labeled "Total" show estimates of the response of total consumption to the first, second, and third rounds.

points in the income distribution is useful for understanding how to best target future payments. Finally, because the survey asks for a wide range of other types of information—such as income, employment, demographics, and wealth—it provides a richer set of potential covariates. This information also allows the authors to identify households in different wealth groups and with different work-from-home potential, which, in turn, facilitates their analysis of the degree to which the payments protected particularly vulnerable households from having to reduce their consumption as a result of economic disruptions from the pandemic.

My discussion will highlight three issues. The first is what to make of the key results in the paper and in what sense they represent "relatively low spending responses," as the authors say in the introduction. Figure 1 summarizes the baseline results, showing the point estimates of the marginal propensity to consume (MPC) out of the different waves of EIPs after six months. For now, focus on the solid bars, which show the results for the first two waves; the results for the third wave are discussed below. The left set of bars show the estimated MPCs for expenditures on nondurable goods and services and the right set of bars show the estimated MPCs for total consumption. For the first two rounds of EIPs, the response of nondurables and services spending is indeed low, at 0.12 to 0.15. But the graph also shows that the response of total consumption to the EIPs is much higher.

For total consumption, the MPC ranges from 0.45 to 0.60. I do not find these estimated MPCs very far out of line with work on direct payments to households in earlier recessions or with the findings from the rest of the literature on the pandemic EIPs—particularly considering much of the latter literature focuses on lower-income households, who would generally be expected to have higher propensities to consume.

The comparison merits flagging because the results for the broader measure of consumption suggest that there was indeed a meaningful spending response to the EIPs over the first six months after they went out, at least for the first two waves of EIPs. More generally, it raises questions about what measure of consumption should be the basis for analysis in this context. Over the decades, a large share of studies of consumer spending at the household level have focused on nondurables and services because many theoretical predictions align best with this subset of consumption. But when evaluating EIPs (as well as other countercyclical support directed at households), there is a good case for focusing on broader consumption measures for gauging the degree of stimulus to the macroeconomy since what matters in that context is the full amount of the payment that goes back into the economy. It is also not clear that ignoring purchases of durable goods is the best way to assess the broader effects on household welfare.

As the authors note, the comparison does imply a shift to durable goods relative to the spending response to direct payments in earlier recessions. This finding is unsurprising given what we saw in aggregate consumption data. Aggregate consumer spending on services ran well below its earlier trend in 2020 and 2021, and aggregate consumer spending on durable goods ran well above its earlier trend. The finding is also what one would expect in a period when many people were limiting spending on high-contact services in order to avoid exposure to the virus and reconfiguring their homes to adapt to remote work.

This shift in consumption underscores the unsurprising point that pandemic-specific factors influenced the nature of the response of households to direct payments from the government in this particular episode. One might then ask whether such results are useful at all for informing the use of such payments as a countercyclical tool in future recessions. Given the energy that has already been put into this area of study (with more likely to come), I hope that these authors and others will give some thought as to how to shed light on what their results might look like in the absence of special pandemic factors. For example, something might be learned from comparing results for groups that might have been more or less likely to alter their consumption response because of the pandemic—such as old people and young people or people living in blue states and red states. It also might be useful to compare results for categories of consumption that are likely to be more or less responsive to payments in pandemic times, following Cooper and Olivei (2021), who define categories of "socially distant sensitive spending."

The second issue that I want to highlight is what I will call the "EIP3 mystery." Returning to figure 1, consider the hatched bars, which show the point estimates for the MPC out of the third wave of EIPs after six months. In contrast to the results for the first two waves (shown by the solid bars), the point estimates for the MPC out of the third wave of EIPs suggest a de minimis response for both spending on nondurable goods and services and total consumption spending. This pattern holds even for low-wealth families, as shown in table 7 of the paper. I find the result to be very surprising.

The first thing that we should ask is whether the EIP3 results are plausible given economic conditions at the time. One supporting narrative would be that Americans—across the income distribution—were just very comfortable financially by early 2021 and therefore not constrained in any way that would lead additional income to spur additional spending. It is correct that, despite enormous job loss in 2020, generous government support meant that many Americans' incomes were as high or higher than they were before the pandemic. Between this support and the fact that the pandemic limited consumption opportunities, many did more saving than usual in 2020 (Aladangady and others 2022). However, in order to explain the much lower estimated MPC for EIP3 than for EIP1 and EIP2, one needs to make the case that Americans were in better financial shape in March 2021 than they were when the two earlier waves were disbursed. I do not see the available data being strongly supportive of that view. Incomes were not particularly robust at that point, with expanded unemployment benefits having ended six months earlier and job postings just beginning to pick up. Various measures of financial stress from the US Census Bureau's Household Pulse Survey, such as the share of respondents reporting difficulty paying their expenses and the share reporting food insecurity, were basically in line with readings over much of the pandemic period. In early March 2021, the JPMorgan Chase Institute's estimates of median checking account balances for the first and second quartiles of the income distribution were about at the average level seen over the pandemic to date (Greig and Deadman 2022).

I also do not see a strong case for Americans being less interested in spending when EIP3 was disbursed than at the time of the two earlier waves. Although households may have stocked up on many types of durable goods over the preceding year, pent-up demand for services was presumably high and the ramping up of vaccination rates was making it much safer to act on that demand.

What about the direct evidence regarding EIP3? Although there is as yet little formal analysis (beyond this study), the informal evidence suggests a material spending response. Table 2 in the paper shows that when CE survey respondents were directly asked how they would use EIP3, more than half said they would use it mostly for expenses—a share that is only a bit lower than that for the earlier two waves. Similarly, Gelman and Stephens (2022) find that the share of Household Pulse Survey respondents reporting they spent out of EIP3 was as large or larger than for the earlier waves. Further, data on credit and debit card spending from the Opportunity Insights Economic Tracker point to a jump in spending after EIP3 was disbursed.[1]

The second question one might ask is what problems there might be with the data or methodology in the paper that could explain why the estimated MPC for EIP3 is so low. The authors flag that the share of CE survey respondents who report not receiving a third payment is much too large (although the degree of underreporting appears to be worse for EIP2) and, in the online appendix, they show some increase in the estimated MPC when they impute what appear to be missing payments. Other possible sources of bias include problems separating the impact of EIP2 from that of EIP3 when the two waves occurred within a couple of months of each other.

All in all, I think the authors are right to downplay their EIP3 results given the concerns about their accuracy. But I also find it dissatisfying that the results and arguments about their validity are so inconclusive and believe it is imperative that researchers dig more deeply to understand the effects of EIP3. The question is important because of the implications for future policy design as well as its relevance to the current debate over inflation. With regard to the latter, there is much speculation that EIP3 helped to spark the sharp increase in inflation that began in spring 2021 by fueling excessive consumer demand (Ngo 2022). But at face value this story does not seem very consistent with extremely low estimated MPC over the six months following the disbursement of EIP3. (Of course, this is not to say that an even greater lagged response could also be contributing to later inflationary pressures.)

Although outside the scope of this exercise, it would be interesting to see if the EIPs might have contributed to the rise in inflation through an effect on labor supply. In particular, did they help fund spells out of the labor

1. Opportunity Insights Economic Tracker, tracktherecovery.org.

force, exacerbating the worker shortage? It does not seem like it would be too hard to explore this question with this data set or a different source.

The third issue I want to highlight is the stimulus role of direct payments to households versus their social insurance role. Macroeconomic textbooks tend to fixate on stimulus as the goal of fiscal measures put in place during recessions. When considering stimulus, the effectiveness of a government spending program is gauged by its multiplier (the amount by which it ultimately raises aggregate demand), which is higher when the MPC is larger. Thus, a finding of a small MPC is sometimes viewed as suggesting a policy was not all that effective.

But stimulus would be an odd primary goal during a pandemic. Encouraging people to spend when a material part of the economy is shut down could be inflationary, and moreover, any type of spending that leads people to get close to each other could foster further spread of the virus. Contrary to what some believe, the arguments against stimulus were largely recognized in the policy community when the spring 2020 COVID-19 relief fiscal packages were put together. (Some experts did make a stimulus-related case that preserving or augmenting spending power would lead to fewer layoffs because businesses would be more confident that demand would be strong once economic activity could safely resume [Blanchard 2020].)

A strength of this paper is that the authors are very clear that stimulus should not be seen as the primary goal of the EIPs. Rather, the authors emphasize the role of the EIPs as social insurance—specifically, in their capacity to prevent hardship associated with current or potential future job loss. And they argue that the higher EIP spending responses for households most likely to experience such hardship—those with low liquidity and low work-from-home potential—demonstrate that the EIPs (at least the first two waves) were successful as social insurance.

I view the social insurance goal as even more expansive than argued by these authors. Specifically, the payments not only had the capacity to prevent immediate hardship in the face of job loss but also were aimed at reducing the likelihood that economic fallout from the COVID-19 pandemic would leave lasting scars on household finances. Such scars were a major consequence of the Great Recession, which (together with the global financial crisis) resulted in deeply weakened household balance sheets for many years. For example, Dettling, Hsu, and Llanes (2018) show that in 2016 wealth for working-age families in the lower 60 percent of the income distribution was still more than 30 percent below the average in 2007. Weak household finances can impair individual household welfare and economic mobility through a variety of channels (Dynan and Wozniak 2021). Also, as discussed by Portes (2020), recession-induced scarring of economic structures can result in slow recoveries and, possibly, permanently lower potential output at the macroeconomic level.

One implication is that research aimed at shedding light on the near-term spending response to EIPs and other countercyclical measures (particularly the response over just a few months) represents only a piece of what is needed to fully assess the benefits of this type of fiscal support for households. There are other important questions to which researchers have not devoted nearly as much attention. For example, were households who experienced unemployment able to maintain higher spending over the long run than those who lost jobs in earlier recessions? Did the EIPs allow households to repay debt or raise their savings on a lasting basis? Did they enable people to quit their jobs and find better ones? Did EIPs facilitate investments like homeownership, starting a business, or postsecondary education?

All in all, I land in the same place as the authors about the social insurance value of the EIPs—they likely provided important protection to economically vulnerable households. In addition, as the authors argue, it would appear that many households that received EIPs were not at particular risk of hardship, suggesting that the same degree of social insurance could have been achieved with less money if the payments were better targeted. In an economy suffering primarily from weak aggregate demand, as would be the case in many recessions, distributing EIPs on a broad basis might still make sense from a stimulus point of view. Finally, while more work needs to be done to assess the contribution that EIPs and other pandemic fiscal support might have made to the sharp rise in inflation since spring 2021, the experience cautions that stimulus measures should be used carefully. Fiscal policymakers need to consider the risk that production will not be able to ramp up as fast as aggregate demand. And monetary policymakers need to be ready to respond if inflation starts to surge.

REFERENCES FOR THE DYNAN COMMENT

Aladangady, Aditya, David Cho, Laura Feiveson, and Eugenio Pinto. 2022. "Excess Savings during the COVID-19 Pandemic." FEDS Notes. Washington: Federal Reserve Board of Governors. https://www.federalreserve.gov/econres/notes/feds-notes/excess-savings-during-the-covid-19-pandemic-20221021.html.

Blanchard, Olivier. 2020. "Designing the Fiscal Response to the COVID-19 Pandemic." Washington: Peterson Institute for International Economics. https://www.piie.com/blogs/realtime-economic-issues-watch/designing-fiscal-response-covid-19-pandemic.

Cooper, Daniel H., and Giovanni P. Olivei. 2021. "High-Frequency Spending Responses to Government Transfer Payments." Working Paper 21-10. Federal Reserve Bank of Boston.

Dettling, Lisa, Joanne Hsu, and Elizabeth Llanes. 2018. "A Wealthless Recovery? Asset Ownership and the Uneven Recovery from the Great Recession." FEDS Notes. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/econres/notes/feds-notes/asset-ownership-and-the-uneven-recovery-from-the-great-recession-20180913.html.

Dynan, Karen, and Abigail Wozniak. 2021. "Family Wealth as an Engine for Macroeconomic Growth." In *The Future of Building Wealth: Brief Essays on the Best Ideas to Build Wealth—for Everyone*, edited by Ray Boshara and Ida Rademacher. Washington: Aspen Institute.

Gelman, Michael, and Melvin Stephens Jr. 2022. "Lessons Learned from Economic Impact Payments during COVID-19." In *Recession Remedies: Lessons Learned from the U.S. Economic Policy Response to COVID-19*, edited by Wendy Edelberg, Louise Sheiner, and David Wessel. Washington: Brookings.

Greig, Fiona, and Erica Deadman. 2022. "Household Pulse: The State of Cash Balances through March 2022." JPMorgan Chase Institute. https://www.jpmorganchase.com/institute/research/household-income-spending/household-pulse-cash-balances-through-march-2022.

Ngo, Madeleine. 2022. "Is the Stimulus to Blame for High Inflation?" *Vox*, October 17. https://www.vox.com/policy-and-politics/2022/10/17/23401726/inflation-stimulus-american-rescue-plan.

Portes, Jonathan. 2020. "The Lasting Scars of the COVID-19 Crisis: Channels and Impacts." *VoxEU*, June 1. London: Centre for Economic Policy Research. https://cepr.org/voxeu/columns/lasting-scars-covid-19-crisis-channels-and-impacts.

COMMENT BY

MATTHEW ROGNLIE The last two decades have witnessed a revolution in how macroeconomists model household savings and consumption. Gone is the representative agent, with its infinite horizon and low marginal propensity to consume (MPC). In its place, we now have households subject to incomplete markets and credit constraints, with shorter effective horizons and much higher MPCs. The macro consequences of this shift are profound: monetary policy works through different channels, and deficit-financed fiscal policy is vastly more powerful.

This revolution has been driven in part by an influential series of empirical papers documenting high MPCs out of unexpected income shocks. Chief among these are two papers studying the 2001 and 2008 stimulus payments in the United States: Johnson, Parker, and Souleles (2006) and Parker and others (2013).

The recent pandemic brought similar payments but at a vastly larger scale: as noted by the authors, the three Economic Impact Payments (EIPs) in 2020–2021 totaled about $800 billion, whereas the 2008 program paid about $120 billion in 2020 dollars (Parker and others 2013), and the 2001 program was smaller still (Johnson, Parker, and Soules 2006). In light of the first two papers' influence, it is only natural to pursue a similar study of the new, far larger payments, and I am delighted these authors—two of whom worked on the first two papers—have taken up the challenge.

And it *is* a challenge, because the key source of identification for previous studies—random variation in the timing of disbursement—is now virtually absent. Instead, the authors must rely on variation in the receipt and amount of EIPs, both of which are nonrandom and determined by variables like income and number of children. If these variables are correlated with fluctuations in consumption that happened for some other reason—quite conceivable in the volatile pandemic environment—then clean identification is in doubt.

The authors, of course, are aware of this challenge and rise to the occasion. Their major conclusion, which I think is quite credible, is that the short-term spending response to the 2020–2021 EIPs was smaller than for the stimulus payments in 2001 and 2008.

One notable aberration is that the authors find seemingly no effect for the third EIP: for broader consumption measures, none of the estimates are statistically significant, and the point estimate on the cumulative two-quarter effect on all Consumer Expenditure (CE) Survey goods and services (table 4) is actually negative. I suspect that this strange result stems from the fact that the effects of the second and third EIPs are not separately identified; the two EIPs happened in short succession and had broadly similar eligibility criteria and phaseout rules. Some of the effect of the third EIP, therefore, is likely being assigned to the second EIP instead, which has a rather high point estimate for the two-quarter overall MPC (0.601).

If we adjust for this issue, however, the paper's core message remains intact: MPCs out of the 2020–2021 payments, though still far too high to be consistent with a permanent income model, were lower than the corresponding MPCs in 2001 and 2008. In the remainder of this discussion, I will explore the macroeconomic implications of this finding. In particular, I ask: If MPCs out of these payments were lower in the first few quarters, does that mean the payments had a smaller effect on aggregate demand? Or was this effect merely delayed? If the latter, perhaps the payments contributed to the surge in excess demand and inflation experienced over the last year and a half.

To help answer these questions, I outline a simple theoretical framework for the dynamics of household consumption following a government transfer. This framework provides several general insights into fiscal transmission—for instance, that excess savings following a transfer dissipate more slowly than a partial equilibrium view would imply, leading to a more persistent output effect. I then perform an experiment where I temporarily decrease MPCs following the transfer, consistent with their apparent decline in the data, and show how this results in a delayed output effect from the transfer. Finally, I discuss two possible deficiencies in my framework: the lack of long-term savings, and the lack of inelastic asset markets. Accounting for the former might decrease the output effect of a transfer, but the latter works in the opposite direction, introducing a new and potentially powerful channel of transmission to aggregate demand.

THEORETICAL FRAMEWORK I now sketch a simple framework for the propagation of fiscal transfers in a population featuring limited heterogeneity, with different household types $i = 1, \dots, N$. This is a discrete-time version of the continuous-time framework in Auclert, Rognlie, and Straub (2023), which has many of the same results, along with some extensions. All variables are in level deviations from steady state.

Assume that if household i's cash on hand in period t—including both assets from the previous period and income this period—increases by x_t, then the household will consume an additional $mpc_i x_t$, where $mpc_i \in [0, 1]$ is some type-specific constant.[1] Households are myopic and do not anticipate that future income or taxes will deviate from steady state. The steady-state real interest rate is $r = 0$, and the central bank sets its policy rate to maintain $r_t = r = 0$ in all periods, neither stimulating nor contracting demand.[2] Nominal wages are rigid, production is linear in labor, and at the margin households are forced to supply extra labor hours to fulfill any increase in demand. As a result, if total goods demand increases from steady state by y_t, the income of each household i increases by $\theta_i y_t$, for some $\theta_i > 0$ satisfying $\sum_{i=1}^{N} \theta_i = 1$.

1. This can be microfounded as the first-order solution to a model with concave utility in assets; see Auclert, Rognlie, and Straub (2023).

2. These assumptions facilitate a pen-and-paper solution of the model. As Auclert, Rognlie, and Straub (2023) show, relaxing them—by introducing rational expectations of income or monetary policy that raises real interest rates in a boom—tends to shrink and shorten the demand effects of a transfer. On the other hand, monetary policy that cuts real interest rates in a boom—for instance, because it is at the zero lower bound and inflation rises—amplifies the demand effects.

Assume further that household type N is Ricardian with $mpc_N = 0$, which is the MPC consistent with a permanent-income household on its Euler equation in the limit $r \to 0$ and $\beta \to 1$. When this household receives additional income, it saves that income forever. All other households, in contrast, are assumed to be non-Ricardian, with $mpc_i > 0$. The Ricardian household can be interpreted either as a wealthy infinite-horizon household or as a proxy for other recipients of marginal spending that are unlikely to spend domestically out of their receipts, such as the government or foreigners.

Finally, coming into period 0, assume that the government makes type-specific transfers (EIPs), which effectively increase the initial asset positions $a_{i,-1}$ of each household type. It rolls over the increased debt from these transfers forever at the real interest rate $r = 0$.

The evolution of this economy away from steady state is summarized by the equations

$$y_t = \sum\nolimits_i mpc_i \left(a_{it-1} + \theta_i y_t\right) \text{ and} \tag{1}$$

$$a_{it} = \left(1 - mpc_i\right)\left(a_{it-1} + \theta_i y_t\right), \tag{2}$$

where, again, both y_t and a_{it} denote deviations from steady state in levels. The increase in cash on hand—assets and income—for household type i is $a_{it-1} + \theta_i y_t$, of which the household consumes mpc_i. Summing these increments to goods demand across all i gives output y_t in equation (1). Equation (2) then gives the evolution of assets: at the end of period t, household type i saves the unconsumed portion of cash on hand as assets a_{it}.

There are several ways to solve for equilibrium in this model. First, we can solve equation (1) for each t sequentially, obtaining

$$y_t = \left(1 - mpc\right)^{-1} \sum\nolimits_i mpc_i a_{it-1}, \tag{3}$$

where we define $mpc \equiv \sum_i \theta_i mpc_i$ to be the average MPC out of marginal income and then plug y_t into equation (2) to obtain assets for the next period. This is a period-by-period Keynesian multiplier, where the impulse $\sum_i mpc_i a_{it-1}$ to spending is amplified by $(1 - mpc)^{-1}$.

Alternatively, we can take $a_{i,-1}$ and the sequence $\{y_t\}$ to be exogenous, iterate on equation (2) to obtain the implied sequence of assets, and then calculate the implied sequence of consumption $c_{it} = mpc_i (a_{it-1} + \theta_i y_t)$. If there is a shock to income $\theta_i y_s$ at date s, then coming into date t, a fraction

$(1 - mpc_i)^{t-s}$ of that income will remain, of which mpc_i will be spent at date t. The matrix $\mathbf{M}_i$ that maps sequences of income $\{\theta_i y_s\}$ to consumption $\{c_{it}\}$ therefore has entries $M_{i,ts} = mpc_i(1 - mpc_i)^{t-s}$ for $t \geq s$ and $M_{i,ts} = 0$ for $t < s$. Aggregating across all households i, the matrix mapping $\{y_s\}$ to $\{c_t\}$ is then $\mathbf{M} \equiv \sum_i \theta_i \mathbf{M}_i$. This is the matrix of intertemporal MPCs introduced by Auclert, Rognlie, and Straub (2018).

Defining $c_{it}^{PE} = M_{i,t0} a_{i,-1}$ to be household i's partial equilibrium consumption response to the fiscal shock—the path of consumption ignoring any changes in aggregate $\{y_t\}$—and aggregating to $c_{it}^{PE} = \sum_i c_{it}^{PE}$, equilibrium output is characterized by an intertemporal Keynesian cross

$$\mathbf{y} = \mathbf{M}\mathbf{y} + \mathbf{c}^{PE}, \tag{4}$$

where $\mathbf{y}$ and $\mathbf{c}^{PE}$ are vectors stacking the sequences $\{y_t\}$ and $\{\mathbf{c}^{PE}\}$. In this case, it turns out that the solution to equation (4) is given by

$$\mathbf{y} = \left(\mathbf{I} + \mathbf{M} + \mathbf{M}^2 + \cdots\right)\mathbf{c}^{PE}, \tag{5}$$

where $\mathbf{I}$ is the identity matrix. This is a direct intertemporal generalization of the traditional Keynesian multiplier process, where $1/(1 - mpc)$ is written $1 + mpc + mpc^2 + \ldots.$

Partial sums in equation (5) can be interpreted as rounds of general equilibrium adjustment. The sequence $\mathbf{c}^{PE}$ alone is partial equilibrium spending; $(\mathbf{I} + \mathbf{M})\mathbf{c}^{PE}$ takes into account that this spending creates additional income, which is spent; $(\mathbf{I} + \mathbf{M} + \mathbf{M}^2)\mathbf{c}^{PE}$ takes into account the income created by that spending; and so on. After infinitely many rounds, this process converges to the general equilibrium $\mathbf{y}$.[3]

Results about equilibrium. We can quickly derive several features of equilibrium, summarized as:

- Result 1: in the long run, the Ricardian household owns all the additional assets.
- Result 2: general equilibrium output $\mathbf{y}$ is greater than partial equilibrium spending $\mathbf{c}^{PE}$, and in the long run y_t decays at a slower rate than c_t^{PE}.

3. For the general case covered in Auclert, Rognlie, and Straub (2018), this iterative process does not necessarily converge to a finite time path. Here, however, convergence is easy to prove, because the existence of Ricardian households $\theta_N > 0$ implies that the l^1 norm of $\mathbf{M}$ is strictly less than one.

- Result 3: the cumulative output effect of the transfer is given by the simple formula:

$$\sum_{t=0}^{\infty} y_t = \theta_N^{-1} \sum_{i=1}^{N-1} a_{i,-1}. \tag{6}$$

How do we derive these results? Result 1 follows from equation (3), which implies that y_t is bounded from below by $(1 - mpc)^{-1}$ $(\min_{i<N} mpc_i)\sum_{i<N} a_{it-1}$. Hence, given total non-Ricardian assets $\sum_{i<N} a_{it-1}$ coming into period t, y_t will be a strictly positive multiple of that, and a share $\theta_R y_t$ will be received by the Ricardian household and immobilized. Over time, this implies an exponential decline in total non-Ricardian assets, which trickle up (Auclert, Rognlie, and Straub 2023) to the zero-MPC Ricardian household. This is in line with empirical evidence showing that poorer households deplete their transfers more quickly than wealthy ones.

The first part of result 2, that $\mathbf{y}$ is larger than $\mathbf{c}^{PE}$, follows directly from equation (5). To understand the second part, note that if all households receive transfers coming into date 0, then c_t^{PE} asymptotically decays at a rate of $1 - \min_{i<N} mpc_i$, corresponding to the non-Ricardian household with the lowest MPC. But in general equilibrium, this household will receive back the income from some of its own spending, and its assets will not decay as quickly.[4] This leads to a more persistent output effect.

Finally, result 3 comes from the fact that all assets transferred to non-Ricardian households must eventually end up in the hands of the Ricardian household. In general equilibrium, this happens via increases in output, but only a fraction θ_N of increased output is earned by the Ricardian household, and hence cumulatively, output needs to increase by θ_N^{-1} times the extra assets held by non-Ricardian households.[5] Remarkably, equation (6) makes

4. Formally, we can condense equations (2)–(3) to get a law of motion $\mathbf{a}_t = (\mathbf{I} - \text{diag}(\mathbf{mpc}))(\mathbf{I} + (1 - mpc)^{-1}\theta\mathbf{mpc}')\mathbf{a}_{t-1}$, where we stack non-Ricardian households $i = 1, \ldots, N - 1$ in bolded vectors. Perron-Frobenius implies that the matrix mapping $\mathbf{a}_{t-1}$ to $\mathbf{a}_t$ has a unique leading positive eigenvalue λ with positive eigenvector $\mathbf{v}$, which governs asymptotic decay. We can write the equation for this eigenvector as $(\lambda - (1 - mpc_i))\, v_i = \dfrac{1 - mpc_i}{1 - mpc}\theta_i \sum_j mpc_j v_j$, and from positivity of $\mathbf{v}$ it follows that $\lambda \geq 1 - mpc_i$ for all i, and indeed that strictly $\lambda > 1 - mpc_i$ if there is any non-Ricardian agent with $mpc_i < 1$.

5. Another interpretation is provided by the formula in equation (5). Multiplying a sequence by the row vector of all ones, $\mathbf{1}'$, takes its sum. One can show that $\mathbf{1}'\mathbf{M}$ equals $(1 - \theta_N)\mathbf{1}'$, since the entire income share $1 - \theta_N$ received by non-Ricardian households is eventually spent. Multiplying equation (5) on the left by $\mathbf{1}'$, it becomes $\mathbf{1}'\mathbf{y} = (1 + (1 - \theta_N) + (1 - \theta_N)^2 + \ldots)\mathbf{1}'\mathbf{c}^{PE} = \theta_N^{-1}1'\mathbf{c}^{PE}$. It is easy to show that $\mathbf{1}'\mathbf{c}^{PE} = \sum_{i=1}^{N-1} a_{i,-1}$, since the cumulative partial equilibrium increase in consumption equals the initial excess assets.

no reference to the MPCs of the non-Ricardian agents: all that matters for the cumulative output effect is that these MPCs are positive, so that any cash received is eventually spent.[6]

APPLYING THE FRAMEWORK

Calibration. Now that the theoretical framework has been established, I will discuss quantification. I consider a case where there are only three household types. First, type 1 is hand-to-mouth, with $mpc_1 = 1$. Second, type 2 has an intermediate $mpc_2 = 0.2$, and I call it a target household since it reverts to its steady state asset target at a rate of 0.2 per quarter. Finally, type 3 is Ricardian, with $mpc_3 = 0$.

My main calibration will feature all three of these types, with $\theta_1 = 0.1$, $\theta_2 = 0.4$, and $\theta_3 = 0.5$. In line with the broader interpretation discussed above, the high Ricardian share is intended to capture marginal recipients of aggregate spending that likely have a low or zero MPC: the government (through taxes), foreigners, some business profits, and a small fraction of labor earnings. If aggregate income increases at date t, these assumptions on θ_i imply an aggregate MPC in the first year, quarters t through $t + 3$, of 0.34, and an aggregate MPC in the second year, quarters $t + 4$ through $t + 7$, of 0.10.

Assuming that only 0.1 out of the $\theta_3 = 0.5$ is earned by labor, we can normalize these intertemporal MPCs by total labor earnings 0.6, obtaining a first-year MPC of 0.56 and a second-year MPC of 0.16. Importantly, these are very close to the first two annual intertemporal MPCs, weighted by labor earnings, reported by Auclert, Rognlie, and Straub (2018) using data from Fagereng, Holm, and Natvik (2021).

Finally, I assume that the transfer is relatively progressive: from the unit transfer, the non-Ricardian households receive a higher share $a_{i,-1}$ than their ordinary share of marginal income θ_i. In particular, $a_{1,-1} = 0.2$ and $a_{2,-1} = 0.6$.

Beyond the main calibration described so far, to better understand mechanisms I will also consider two related calibrations, both of which have only one non-Ricardian household: an "only hand-to-mouth" calibration where $\theta_1 = 0.5$, $a_{1,-1} = 0.8$, and $\theta_2 = a_{2,-1} = 0$; and an "only target" calibration where $\theta_2 = 0.5$, $a_{2,-1} = 0.8$, and $\theta_1 = a_{1,-1} = 0$. Note that in all these cases, since the allocation of both the transfer and marginal income between non-Ricardian and Ricardian households is the same, the cumulative output effect implied by equation (6) is identical.

6. Importantly, however, this result is sensitive to the assumption that the central bank holds the real rate r_t fixed. A rise in r_t provides another mechanism for moving assets from the non-Ricardian households to the Ricardian household, since the latter will generally increase net savings by more in response.

Results. The three panels of figure 1 show the general equilibrium path of output y in the hand-to-mouth, target, and main calibrations. They also show the rounds of adjustment in equation (5) that converge to y: the partial equilibrium round 0 $\mathbf{c}^{PE}$, round 1 $(\mathbf{I} + \mathbf{M})\mathbf{c}^{PE}$, and round 2 $(\mathbf{I} + \mathbf{M} + \mathbf{M}^2)\mathbf{c}^{PE}$. Output $\mathbf{y}$ itself can be viewed as round ∞, since it is the sum $(\mathbf{I} + \mathbf{M} + \mathbf{M}^2 + \ldots)\mathbf{c}^{PE}$.

Despite identical cumulative output effects (result 3), the three calibrations are strikingly different, with impact multipliers varying by a factor of nine. In the hand-to-mouth calibration, the entire output response happens at $t = 0$, as hand-to-mouth households immediately spend both the transfer and the income from the resulting boom, and the excess assets immediately pass to the Ricardian household. In the target calibration, we see the opposite: households slowly draw down their assets, as their increased spending is partly offset by the general equilibrium increase in income, so that assets and spending are more persistent in general than partial equilibrium (result 2). Only a small fraction (about one-ninth) of the cumulative output effect happens on impact.

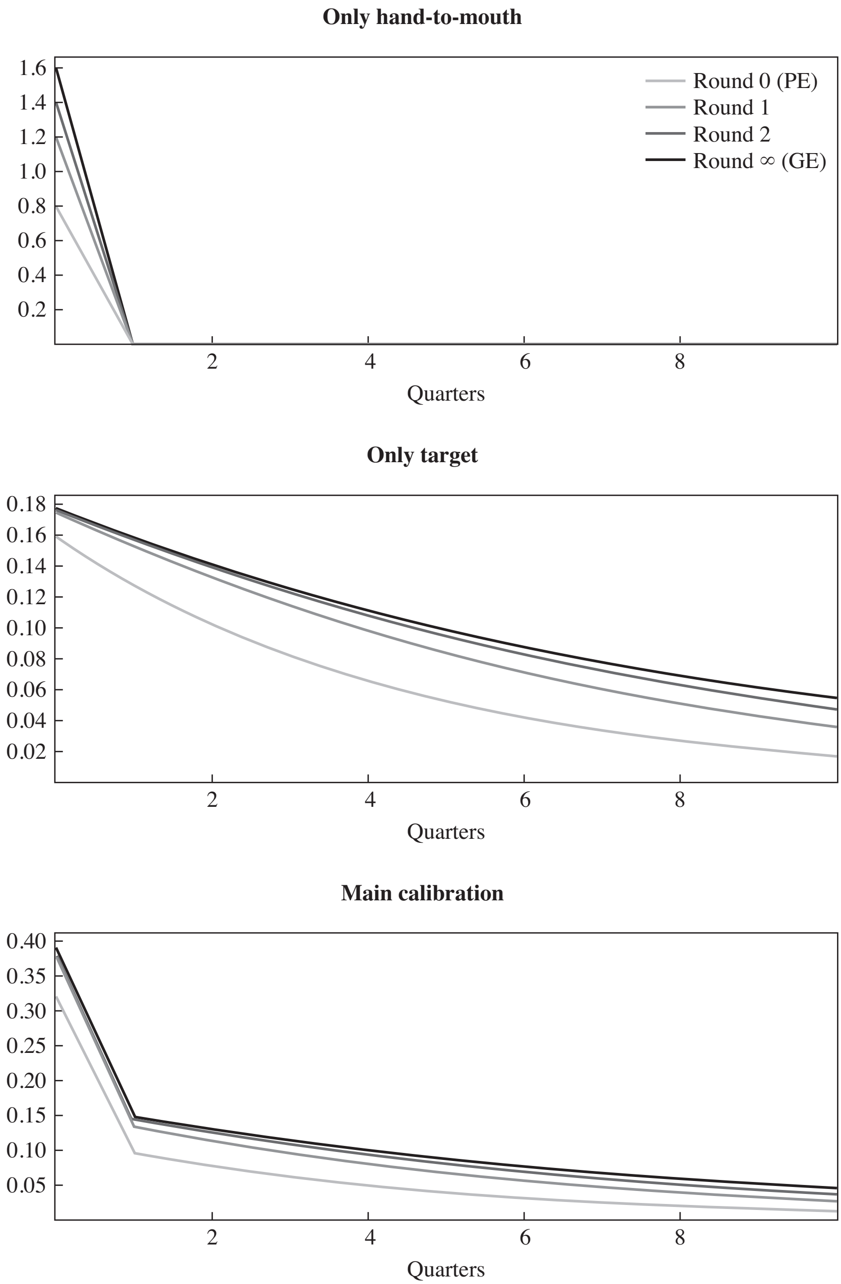
The main calibration, blending hand-to-mouth and target households, is intermediate between these two cases. Thanks to the hand-to-mouth households, there is a spike in output in the quarter of the transfer. But this is still less than one-fourth of the cumulative output effect, which has much higher persistence in general than in partial equilibrium.

The first two panels of figure 2 show the evolution of assets for the main calibration, in both partial and general equilibrium.[7] In the partial equilibrium case, the hand-to-mouth households immediately deplete their assets, and the target households do so at a steady pace, with the vast majority gone after ten quarters. The Ricardian households simply hold on to their initial receipts. In general equilibrium, the hand-to-mouth households still immediately deplete their assets, but the target households do so more slowly, with almost two-thirds of their initial assets remaining after four quarters, and one-third remaining after ten quarters. Total assets remain constant, as assets drawn down by others trickle up to the Ricardian households (result 1).

Experiment: temporarily lower MPCs. As discussed earlier, the evidence from Parker, Schild, Erhard, and Johnson suggests that MPCs out of fiscal transfers may have fallen during the pandemic. This could be due to pandemic-specific circumstances (limited opportunities to spend), nonlinearities in the consumption function (with high liquidity from transfers

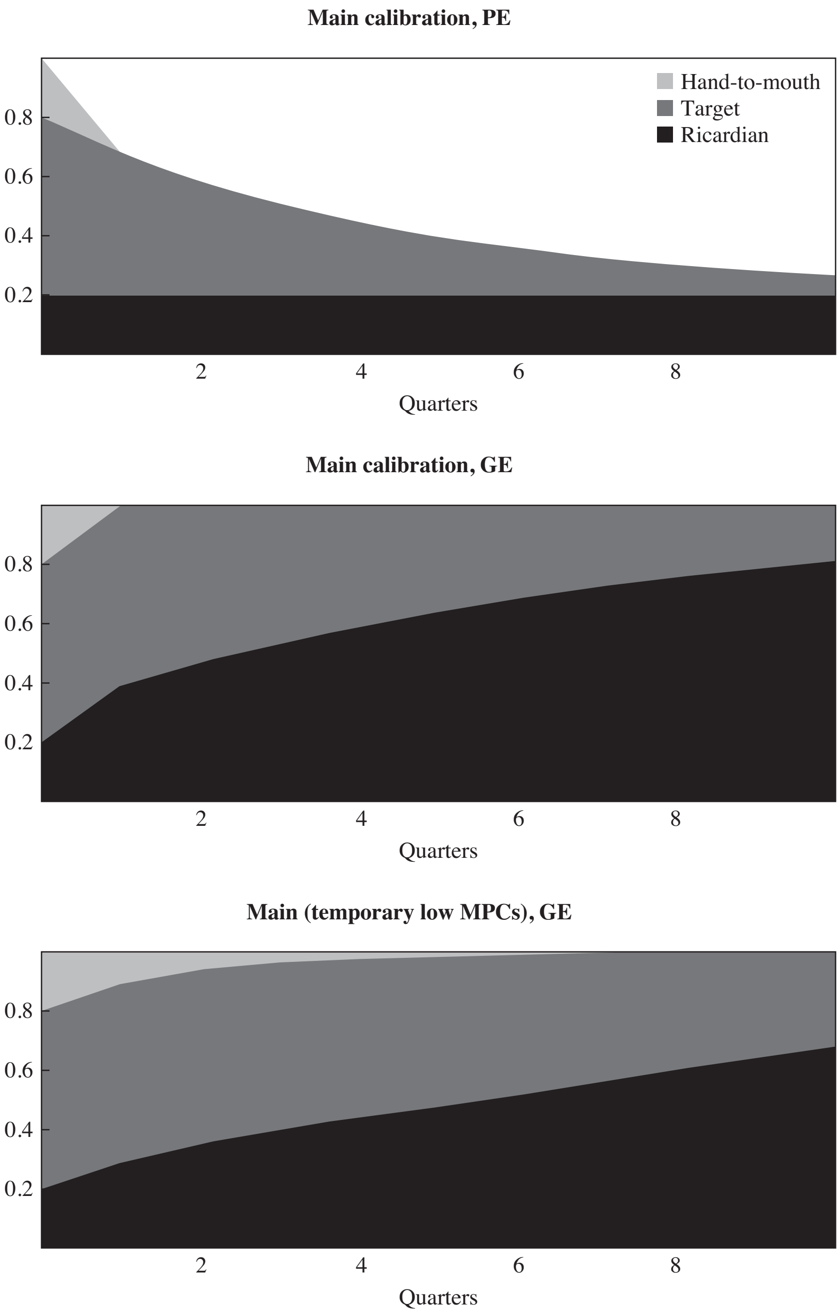
7. At each t, we plot beginning-of-period assets $a_{i,t-1}$ rather than end-of-period assets $a_{i,t}$, so that the initial transfer is visible at $t = 0$.

Figure 1. Output Response to Transfer by Household Calibration and General Equilibrium Rounds



Source: Author's calculations.

Figure 2. Distribution of Assets across Household Types



Source: Author's calculations.

temporarily depressing MPCs), or both. In either case, it seems unlikely that the decline in MPCs is permanent.

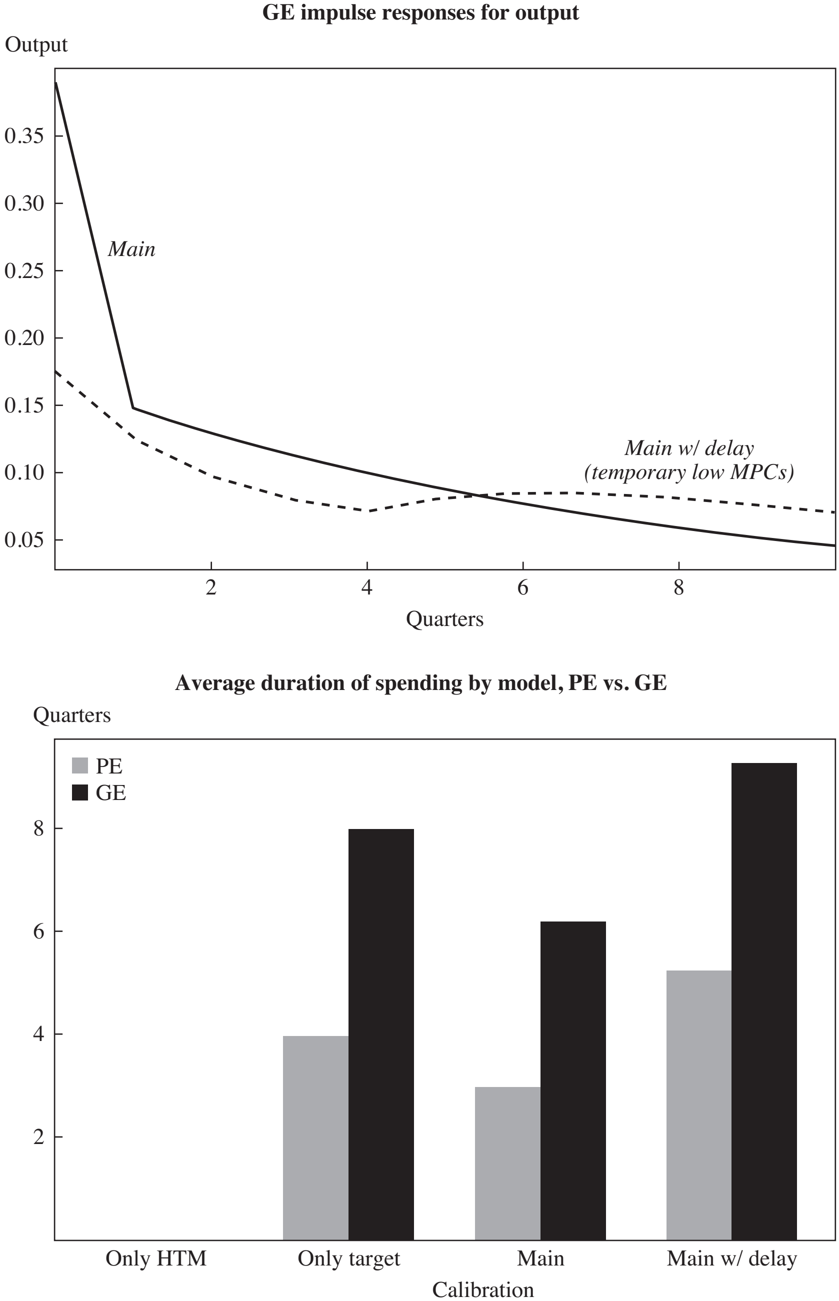
In this experiment, I take a reduced-form approach to think about the effects of declining MPCs. I alter the framework from above by assuming that the MPCs out of excess cash on hand temporarily fall for both hand-to-mouth and target households to half their usual levels, $mpc_{1t} = 0.5$ and $mpc_{2t} = 0.1$ for $t = 0, \ldots, 4$. I assume that these MPCs then converge back to their original levels at a rate of 25 percent per quarter; for example, that $mpc_{1t} = 1 - 0.5 \cdot (0.75)^{t-4}$ for $t \geq 4$. The main calibration is otherwise left unchanged.

The top panel of figure 3 shows the resulting path of output for this delayed spending variant of the model (dashed line), contrasting with the original results (solid line). The impact effect on output, although substantial, is less than half as large, and the path of output is non-monotonic, increasing slightly with the recovery in MPCs after four quarters. Crucially, the cumulative output effect remains the same in both cases (result 3), so that the model with temporarily low MPCs actually has a higher output effect after six quarters, with the gap becoming substantial after eight—making up for the smaller impact effect. The bottom panel of figure 2 shows the corresponding evolution of assets: due to the temporary decline in MPCs, less trickling up of assets takes place than in the original calibration, so that more assets remain with hand-to-mouth and target households, ready to be spent.

Finally, the bottom panel of figure 3 shows the duration of the output increase (or, in partial equilibrium, the increase in household spending) by calibration: the average date at which the increase in output or spending takes place. Across the board, duration is higher in general than partial equilibrium, in line with result 2. Among the original calibrations, it is highest with only target households, and lowest (zero) with only hand-to-mouth households, with the main calibration being in the middle. But the temporary fall in MPCs pushes up duration substantially, to the point where it exceeds every original calibration. Importantly, in all these cases, cumulative output is the same: higher duration simply means that the same overall increase in output is pushed toward later dates.

I suspect that the events of the last few years resemble the delayed-spending case. Although a vast fiscal intervention pushed household liquidity to unprecedented levels, the demand-side effects—though substantial—were not as large as we would normally expect, because MPCs were lower than usual during the pandemic. But since households still had these excess savings on their balance sheets, this merely set us up for a more prolonged

Figure 3. Impulse Responses and Duration across Different Scenarios



Source: Author's calculations.

boom in demand—an inflationary boom that, as of the end of 2022, has not yet receded.

A LINGERING QUESTION FOR FUTURE WORK: THE ROLE OF ASSET MARKETS The framework I have outlined, although useful, relies on one precarious assumption: that whatever portion of a transfer is not consumed by household i today is still subject to the same marginal propensity to consume, mpc_i, in the next period. One can imagine the opposite assumption: that whenever a household receives income, it either consumes that income immediately, or it places the income into long-term savings, out of which the MPC is very low.

In its extreme form, this alternative assumption seems inconsistent with the evidence on intertemporal MPCs highlighted by Auclert, Rognlie, and Straub (2018), which shows that elevated consumption persists for several years following an income shock. (Indeed, I tried to match this evidence in my calibration here.) But that same evidence does allow for some diversion to long-term savings. Indeed, Fagereng, Holm, and Natvik (2021) find that five years after an unexpected income shock, about 10 percent of the income remains unconsumed, and much of this is held in investments like stocks, bonds, and mutual funds.

What if the counterpart of lower MPCs during the pandemic was a much higher allocation to long-term savings? If so, my analysis above would be wrong: it assumes that non-Ricardian households eventually return to their typical high MPCs out of excess assets. If these assets were instead moved to some form of sticky long-term savings, that might never have happened—and the pandemic's low MPCs might have truly dampened the demand effect from transfers, rather than merely delaying it.

But this raises another question: What vehicles were households saving in, and might those have demand effects in their own right? In a simple model where different assets are highly (perhaps perfectly) substitutable, the answer is no: the high substitutability across assets means that the exact choice of where to save is fungible, and in equilibrium it matters little for aggregate outcomes whether a given household invests in stocks, bonds, or deposits. If, however, we assume inelastic markets, in the spirit of Gabaix and Koijen (2021), this changes. Investing in stocks will push up stock prices, potentially leading other households to increase their consumption due to wealth effects, and also to higher corporate investment spending. Investing in real estate will push up real estate prices, allowing existing owners to lever up and increasing both consumption and construction spending. Even a transfer that is saved, if it is saved in the right places, can push up aggregate demand.

At least superficially, this story seems to fit the pandemic experience: as households flush with cash moved into the stock market and real estate—a process already documented in some papers—prices in both markets surged from late 2020 through 2021. This surge in prices likely contributed to aggregate demand and inflation.

Together with Adrien Auclert, Ludwig Straub, and Lingxuan Wu, in ongoing research I am building a theoretical framework to understand this interaction between inelastic markets and aggregate demand. But a great deal of empirical work is also needed. Perhaps the successors to this paper can document not only the marginal propensity to consume, but also the marginal propensity to save in each kind of asset.

REFERENCES FOR THE ROGNLIE COMMENT

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2018. "The Intertemporal Keynesian Cross." Working Paper 25020. Cambridge, Mass.: National Bureau of Economic Research.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2023. "The Trickling Up of Excess Savings." Working Paper 30900. Cambridge, Mass.: National Bureau of Economic Research.

Fagereng, Andreas, Martin B. Holm, and Gisle J. Natvik. 2021. "MPC Heterogeneity and Household Balance Sheets." *American Economic Journal: Macroeconomics* 13, no. 4: 1–54.

Gabaix, Xavier, and Ralph S. J. Koijen. 2021. "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis." Working Paper 28967. Cambridge, Mass.: National Bureau of Economic Research.

Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles. 2006. "Household Expenditure and the Income Tax Rebates of 2001." *American Economic Review* 96, no. 5: 1589–610.

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

GENERAL DISCUSSION Jason Furman remarked that in nominal terms, personal consumption expenditures were more than 10 percent higher in 2021 than in 2019, a sum of $1.5 trillion.[1] This, he believed, was a shocking amount of personal consumption expenditure given the pandemic-induced constraints on services and unemployment levels. He contemplated whether the spike in personal consumption would have happened absent the nearly

1. FRED, "Personal Consumption Expenditures," https://fred.stlouisfed.org/series/PCE#0.

$5 trillion of interventions—maybe, for example, the marginal propensity to consume (MPC) presented in this paper was somehow delayed.[2] He thought it was difficult to explain the elevated level of personal consumption expenditures without some meaningful multiplier on some part of that $5 trillion.

Steven Braun commented that it is still to be seen whether excess savers from 2020 and 2021 will spend their money in 2022. Robert Gordon agreed with Braun, noting that there was a large amount of excess savings. He provided three pieces of evidence on this point. First, he noted that there was a striking upward jump in personal disposable income and savings at the time of the transfers. Second, he stated that excess savings have risen considerably. Compared to the 2019 rate of 7.6 percent of personal disposable income, the value of personal savings increased to $2.4 trillion in mid-2021.[3] Based on Matthew Rognlie's analysis, he interpreted these data as indicating that savings have been gradually shifting from the short-run adjustment agents to the long-run adjustment agents. Third, bank balances have increased by $4.7 trillion (although, he noted, they are not growing as quickly as they did during the period of transfers).[4] He concluded his point by stating that once this liquidity is created, it doesn't go away. It's just shifting from all the people who got the stimulus to the people who saved. William Gale asked if there were data on whether recipients gave money from their Economic Impact Payments (EIPs) to family members, since the options recipients had were to save the money, spend it, or give it away.

Wendy Edelberg also discussed reconciling macroeconomic savings data with the MPCs found in the paper. She agreed with the general principle that if a stimulus payment doesn't show up in consumer spending, it must flow into some type of saving: either paying off debt, deposited in a checking account, or used to buy assets. However, she claimed, the data do not match this prediction. She noted that while there was a big inflow into deposits in the first quarter of 2021, this inflow subsequently stopped. At the same time, there were increases in consumer debt in 2021. Furthermore, a lot of excess savings observed in 2021 came from higher-income people reducing their spending, rather than from lower-income people reducing their spending from income, which at this point also included the EIPs. The MPCs presented in the paper would suggest that a lot of money went

2. Christina D. Romer, "The Fiscal Policy Response to the Pandemic," *Brookings Papers on Economic Activity* (Spring 2021): 89–109.

3. Bureau of Economic Analysis, "National Income and Product Accounts," table 2.6, https://apps.bea.gov/iTable/?reqid=19&step=2&isuri=1&categories=survey.

4. FRED, "Deposits, All Commercial Banks," https://fred.stlouisfed.org/graph/?g=U0OX.

into savings, but Edelberg did not see evidence of this phenomenon in the saving data.

Louise Sheiner was surprised that those reliant on income from jobs with tasks that could not be performed at home had higher MPCs than those with jobs that could be worked remotely. Since a significant fraction of these individuals were unemployed, many were receiving substantial pandemic unemployment insurance payments.

Austan Goolsbee noted that if MPCs are this low, then the immediate impact of the American Rescue Plan should not have been that inflationary because people were actually saving the money. He asked why, then, is there currently so much inflation?

Pierre-Olivier Gourinchas answered by looking at the transfer multiplier, that is, how much aggregate output $1 of fiscal transfer to households causes. He noted that there is a body of work that shows that in a situation characterized by supply constraints and with the assumption of regular MPCs, the transfer multiplier is expected to be very low. Gourinchas reported that the pandemic economy had upward of 70 percent of sectors with supply constraints at one point. His recent research with coauthors found low transfer multipliers—on the order of six cents on the dollar—even with reasonable estimates of the MPC.[5] He concluded that if nominal spending went into supply-constrained sectors, then it was not contributing to real economic activity; instead, it was contributing to inflation.

On this topic, Deborah Lucas expressed her belief that the MPC on the EIPs would have turned out even lower if one also took into account forbearance on student loans and interest payments. The magnitude of the loans especially, she noted, were of similar size to the EIPs.

Claudia Sahm noted that the paper under discussion has a different statistical methodology from the studies done in 2001 and 2008 by Jonathan Parker, David Johnson, and colleagues.[6] That methodology was considered a gold standard because of its quasi-random timing of check disbursement based on Social Security numbers. The current study did not have that same

5. Pierre-Olivier Gourinchas, Şebnem Kalemli-Özcan, Veronika Penciakova, and Nick Sander, "Fiscal Policy in the Age of COVID: Does It 'Get in All of the Cracks'?," in *Economic Policy Symposium Proceedings: Macroeconomic Policy in an Uneven Economy* (Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City, 2021).

6. David S. Johnson, Jonathan A. Parker, and Nicholas S. Souleles, "Household Expenditure and the Income Tax Rebates of 2001," *American Economic Review* 96, no. 5 (2006): 1589–610; Jonathan A. Parker, Nicholas S. Souleles, David S. Johnson, and Robert McClelland, "Consumer Spending and the Economic Stimulus Payments of 2008," *American Economic Review* 103, no. 6 (2013): 2530–53.

element of exogeneity. Sahm pointed to her study with Matthew Shapiro and Joel Slemrod on the Michigan Survey, which does have consistent identification across groups, where they asked questions about the CARES checks.[7] They found monthly spending percentages roughly similar to previous work. Despite certain pandemic-related factors that may have an impact on spending percentages, such as the tendency for people to spend less on vacations and more on food, she said that the spending percentages tend to have time consistency, and that this paper is not consistent with other papers.

Gourinchas discussed why the MPCs presented in the paper may have been lower than expected. If the EIPs functioned as perfect insurance, then the consumption change would be equivalent between those receiving insurance and those not receiving insurance, therefore creating a zero coefficient on the stimulus. If the evidence points to coefficients that are lower than in normal times, it might actually indicate that payments are going in the right direction.

On the point of insurance, Sahm did not believe that MPCs and the speed of spending alone are good measures of insurance. These measurements should be considered alongside many observations to conclude whether the EIP program was effective.

Arvind Krishnamurthy introduced some additional information about the insurance value of the EIPs, noting that household balance sheets and the financial sector did not show evidence of scarring, unlike in 2008. In spring 2020, the prices of securities that were linked to consumer defaults—such as credit card asset-backed securities and loan asset-backed securities—plummeted very quickly. However, these asset prices and their spreads soon after returned to normal levels, which Krishnamurthy read as a sign that households were continuing to service their debts, unlike in 2008.[8]

Caroline Hoxby put forth the idea that the EIP targeting was inexcusably bad, especially for the second and third rounds, since the administrative emergency that was active during the first round of EIP had waned. She thought that while the payments had not done a good job at mitigating losses in the consumption of nondurables, they may have encouraged purchases of nondurables such as the technology, furniture, and other items that allow a person to set up a home office. In this sense, although the

7. Claudia Sahm, Matthew Shapiro, and Joel Slemrod, "Consumer Response to the Coronavirus Stimulus Programs," slides, November 11, 2020, https://drive.google.com/file/d/1zkMXfn4SQMW1mlWTfuEXM-ZXA6Nse0jR/view.

8. Markus Brunnermeier and Arvind Krishnamurthy, "Corporate Debt Overhang and Credit Policy," *Brookings Papers on Economic Activity* (Summer 2020): 447–88.

EIPs may not have worked as intended, the payments may have ensured higher productivity during the pandemic by encouraging consumption of durable goods that improved productivity (for example, home office setup). She believed that the different types of purchases should be identifiable in the data. For instance, durables such as washing machines and cars may not have increased productivity, but electronics, computers, and furniture likely had some effect. This difference could also be isolated by stratifying data between remote and in-person workers.

Gordon agreed with this idea: his research shows that all of the productivity gains since the start of the pandemic were concentrated in the 35 percent of the economy where people primarily work from home. The remaining 65 percent of the economy has negative and zero productivity gains.[9]

Justin Wolfers contended that what appeared to be durables consumption may have followed a pattern more similar to regular consumption of nondurables—as the pandemic began to recede, people's durable purchases (such as desks and treadmills) may have no longer been used. Durables are traditionally thought as goods that yield an ongoing flow of services for many years, but many of the products bought during the pandemic have fallen out of use.

Hoxby also agreed with Rognlie's point that it is possible that the third round of EIP was conflated with the second round. Given their closeness in time, she does not believe there is a good way to separately identify them, as a matter of econometrics.

Sheiner noted that one of the benefits of the EIPs was the fact that they went out quickly, providing temporary liquidity to recipients who were unable to immediately benefit from unemployment insurance.

Adjacent to the insurance topic, Wolfers brought up the proposition that the EIPs may have not only served as insurance but also had income effects on people's labor–leisure decisions, allowing people to stay at home. He thought that evaluating the effect of EIPs on labor supply was a feasible related line of research.

Ayşegül Şahin built off Wolfers's comment on the labor–leisure choice, adding in some reasons for optimism about the EIPs. Şahin noted that Americans' desires to work are declining.[10] At the same time, fewer Americans

9. Robert J. Gordon and Hassan Sayed, "A New Interpretation of Productivity Growth Dynamics in the Pre-pandemic and Pandemic Era U.S. Economy, 1950–2022," working paper 30267 (Cambridge, Mass.: National Bureau of Economic Research, 2022).

10. R. Jason Faberman, Andreas I. Mueller, and Ayşegül Şahin, "Has the Willingness to Work Fallen during the Covid Pandemic?" working paper 29784 (Cambridge, Mass.: National Bureau of Economic Research, 2022).

feel overworked and fewer Americans feel underworked. This evidence suggests that the United States has moved toward a point where work-life balance has improved compared to before the pandemic.

Gale reflected upon some possible areas for further study. He wondered how much impact the EIPs may have had on mental health in relieving pandemic-related anxiety. Braun wondered whether Ricardian equivalence may be influencing EIP recipients' behavior—those who needed the money spent it, while those who didn't need it know they'll be taxed in the future to cover the pandemic spending.

Jonathan Parker addressed some of the topics raised by the discussants in his final remarks. He noted that the goal of the paper was to measure a rapid response to the EIPs. There is substantial weakening of statistical power when any lag responses are measured, so he and his coauthors did not try to make the stronger claim that they measured these lagged responses. The low MPCs that the authors found are not an argument that the EIPs were not spent at all. Rather, they were a piece of evidence that points to the EIPs not being spent immediately. He emphasized Rognlie's model that showed how various agents may spend their payments over short and long periods of time. The authors' intentions were to see whether the payments were being spent rapidly and timely as pandemic insurance.

ZHENGYANG JIANG
Northwestern University Kellogg School of Management

HANNO LUSTIG
Stanford Graduate School of Business

STIJN VAN NIEUWERBURGH
Columbia Business School

MINDY Z. XIAOLAN
University of Texas at Austin McCombs School of Business

Measuring US Fiscal Capacity Using Discounted Cash Flow Analysis

ABSTRACT We use discounted cash flow analysis to measure the projected fiscal capacity of the US federal government. We apply our valuation method to the Congressional Budget Office (CBO) projections for the US federal government's primary deficits between 2022 and 2052 and projected debt outstanding in 2052. The discount rate for projected cash flows and future debt must include a GDP or market risk premium in recognition of the risk associated with future surpluses. Despite current low interest rates, we find that US fiscal capacity is more limited than commonly thought. Because of the back-loading of projected primary surpluses, the duration of the surplus claim far exceeds the duration of the outstanding Treasury portfolio. This duration mismatch exposes the government to the risk of rising interest rates, which would trigger the need for higher tax revenue or lower spending. Reducing this risk by front-loading primary surpluses requires a major fiscal adjustment.

Recently, there has been an active debate about the fiscal capacity of the United States and other countries, but there is no consensus on the proper measurement of fiscal capacity. Some economists have argued that we can use the ratio of the government's interest expense over GDP as a

Conflict of Interest Disclosure: This research was supported by the National Science Foundation grant no. 2049260. Stijn Van Nieuwerburgh is a director at Moody's Investor Services. 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. Other than the aforementioned, the authors are not currently an officer, director, or board member of any organization with a financial or political interest in this paper.

Brookings Papers on Economic Activity, Fall 2022: 157–209

measure of fiscal capacity (Furman and Summers 2020). Others have argued that we should compare the risk-free rate to the growth rate of the economy (Blanchard 2019; Andolfatto 2020). Most authors have concluded that low interest rates have substantially increased US fiscal capacity.

We define a country's projected fiscal capacity as the present discounted value (PDV) of that country's projected future primary surpluses. We apply our method using the long-term budget projections from the Congressional Budget Office (CBO) as the point estimate of future cash flows.

In standard models with long-horizon investors, the government's debt is fully backed by future surpluses. The measurement of fiscal capacity then becomes a forward-looking valuation exercise. The country's actual fiscal capacity can differ from our projected measure if the market's valuation of the debt exceeds the PDV of projected surpluses. This means that the market is pricing in a large fiscal correction relative to the projections.

Our definition of fiscal capacity differs from the one commonly used by the International Monetary Fund (IMF), the Organisation for Economic Co-operation and Development (OECD), and other institutions that use a marginal definition of fiscal capacity or fiscal space: the ability to issue additional debt in response to a shock (Botev, Fournier, and Mourougane 2016). These distinct concepts are connected. If a country's projected fiscal capacity is low relative to the value of debt outstanding, then the country's ability to issue additional debt at low interest rates is impaired. Our approach has a number of advantages. First, our measure is easily quantifiable. We come up with a dollar amount, not a combination of indicators as produced by the IMF or OECD. Second, our approach is founded in modern finance. We rule out free lunches for the government and apply textbook finance to the Treasury's balance sheet.[1]

We propose a simple, easy-to-implement discounted cash flow approach. As in any valuation exercise, this approach requires estimating the discount rate as well as forecasting the underlying cash flows, tax revenues minus non-interest government spending. A proper discount rate for projected surpluses and future debt must reflect the riskiness of the underlying cash flows. Following our previous work (Jiang and others 2019), we develop an upper bound on fiscal capacity by using the expected return on a claim to GDP, also known as the total wealth or market portfolio, to discount future taxes, spending, and future debt. This approach implicitly assumes that surpluses are as risky as GDP. That is a conservative assumption because

1. See Lucas (2012) on the importance of proper risk adjustment when evaluating government policies.

the surplus/GDP ratio is pro-cyclical. As a result, our risk adjustment is too small and our measure of projected fiscal capacity will tend to overstate actual fiscal capacity.

This discount rate is the sum of a maturity-specific risk-free interest rate and the GDP risk premium. We argue that a plausible value for the GDP risk premium should be at least 2.5 percent per year. When we use this discount rate, the PDV of future debt is well-behaved even when the risk-free rate is lower than the growth rate.

In the discounted cash flow approach, the PDV of debt outstanding in the distant future converges to zero. The transversality condition (TVC) holds because the discount rate applied to future debt includes a GDP risk premium.[2] Hence our definition of fiscal capacity as the PDV of future primary surpluses. We explore a conservative scenario in which the debt projected by the CBO at the end of the projection horizon is fully backed by subsequent surpluses.

In spite of the secular decline in long rates, we find that US fiscal capacity is limited. The CBO projects average primary surpluses of −3.2 percent of GDP between 2022 and 2052. The PDV of these projected surpluses is −\$21.16 trillion in 2021 dollars. In addition, the CBO-projected debt outstanding in 2052 is 185 percent of GDP. Starting in 2053, the United States would need to generate a steady-state surplus of 2.16 percent to pay back the debt outstanding in 2052. Discounted back to 2021 at the appropriate discount rate, the 2052 debt is worth about \$33.5 trillion. When we combine the PDV of projected surpluses until 2052 of −\$21.16 trillion with the PDV of the projected debt outstanding in 2052 of \$33.5 trillion, we end up with an upper bound on the projected fiscal capacity of \$12.34 trillion. This is our baseline estimate of (an upper bound on) projected fiscal capacity. It falls about \$10 trillion short of the actual \$22.28 trillion value of all US Treasuries outstanding at the end of 2021. This gap occurs even though we assumed a large, permanent fiscal (primary surplus) correction of 5.36 percent of GDP per year after 2055 relative to the 2022–2052 period.

Alternatively, instead of using the CBO-projected debt in 2052, we can back out the annual surplus after 2052 that is required to match the value of outstanding debt at the end of 2021 to the PDV of all surpluses after 2021. We find that the United States would need to run a permanent primary surplus of 2.79 percent of GDP after 2052, a 5.98 percent fiscal correction relative to the pre-2052 path for surpluses.

2. In the discounted cash flow approach, the TVC only fails if the GDP risk premium is smaller than the difference between the growth rate of the economy and the risk-free interest rate. This is not the case for the United States.

An extended measure of the projected fiscal capacity includes the seigniorage revenue earned by the Treasury. US Treasuries earn a convenience yield because they play a special role in the global financial system. Adding the present value of these seigniorage revenues of $4.04 trillion brings our final estimate for the upper bound on fiscal capacity in 2021 to $16.42 trillion, or 73 percent of 2021 GDP. This estimate remains substantially below the observed value of debt to GDP at the end of 2021. Despite the current low interest rates (and hence low debt service), and even after considering the special status of Treasuries, we find that US fiscal capacity is quite limited.

There are three potential explanations for the large gap between the PDV of surpluses we compute and the market's valuation of Treasuries. First, the Treasury market is right and rational market participants anticipate either a large fiscal correction that is not reflected in the CBO's projections or Japan-style financial repression. We quantify this correction in the paper. Second, the market is wrong. Investors—as of the end of 2021—could be overly optimistic about future surpluses or fail to price in future inflation. Mispricing in financial markets can persist for long periods of time. Third, the market may anticipate a switch from the current regime with pro-cyclical primary surpluses to one with countercyclical surpluses. Making tax revenues countercyclical would lower the discount rate on the tax revenue claim, and making non-interest spending pro-cyclical would increase the discount rate on the spending claim. Section V shows that this change is the most potent way of boosting projected fiscal capacity. It is arguably also the most painful and hence least politically feasible way, since it requires belt tightening at the worst possible (high marginal utility) times.

Projected fiscal capacity may be even more limited than what our calculations suggest for three reasons. First, our estimates put only an upper bound on fiscal capacity because we assume that surpluses are only as risky as GDP. Second, our estimates of fiscal capacity assume that a major fiscal adjustment will take place after 2052, turning from large primary deficits to large primary surpluses. This is an adjustment unlike any other in US history. Third, the GDP risk premium estimate used in our discount rate is at the lower end of the empirically plausible range. Each of these assumptions, discussed in detail below, increases our estimate of projected fiscal capacity and shrinks the gap with the observed debt/GDP ratio. Each one makes our calculations conservative and reinforces our conclusion that fiscal capacity is more limited than commonly thought.

Typically, a pension fund will seek to match the duration of the cash inflows from its portfolio to the duration of the cash outflows to its retirees to avoid interest rate risk. The US Treasury has not matched the duration of its projected cash inflows, primary surpluses, to the duration of its outflows, coupon and principal payments on the debt portfolio. Because of the back-loading of projected primary surpluses, the duration of its asset claim is very long (283 years in the baseline model), much longer than the duration of its outstanding bond portfolio (around five years in 2021). This creates a large duration mismatch. When rates increase, US fiscal capacity, the present value of future surpluses, decreases dramatically, but the value of its liabilities, the portfolio of outstanding Treasury debt, decreases by much less. As a result, an interest rate increase will require large fiscal adjustments. A mere 1 percentage point increase in yields of all maturities, holding constant nominal GDP growth and projected primary surpluses until 2052, requires an increase in surpluses of 2.67 percent of GDP each year after 2052 relative to the baseline model.

The large realized changes in interest rates between December 31, 2021, and May 31, 2022, when interest rates moved up anywhere from 130 to 175 basis points along the term structure, are a concrete example of this duration argument. These changes require a massive increase in primary surpluses after 2052 to maintain the same projected fiscal capacity: from 2.16 percent per year to 6.24 percent per year.

From an optimal maturity management perspective, that is, to avoid costly variation in tax rates, the Treasury should either front-load surpluses to shorten the duration of its assets or increase the maturity of its outstanding debt or both. In order to eliminate the duration mismatch completely, we find that the Treasury would have to increase the primary surplus by 6 percent of GDP each year between 2022 and 2052, relative to the CBO's baseline projections.

We develop intuition for these quantitative estimates by examining the steady state in which the surplus is a constant share of GDP. In the steady state, fiscal capacity relative to GDP equals the price/dividend ratio on the GDP claim multiplied by the steady-state surplus/GDP ratio. The price/dividend ratio determines the fiscal capacity per dollar of surplus, expressed as a percentage of GDP. This price/dividend ratio depends on the risk-free rate, the term premium, the GDP risk premium, and the expected growth rate of GDP. An increase in the expected growth rate, a decrease in the risk-free rate, a decrease in the term premium, or a decrease in the GDP risk premium all increase fiscal capacity.

We estimate the price/dividend ratio for the total wealth portfolio to be around 86 at the end of 2021, which implies an estimate for total wealth, including human wealth, of about eighty-six times GDP. To get an upper bound on the fiscal capacity of 99.6 percent of GDP, the size of the debt/GDP ratio at the end of 2021, the United States would need a steady-state primary surplus of 1.16 percent. Relative to the aforementioned CBO projections of average primary surpluses of −3.2 percent between 2022 and 2052, this requires a major fiscal correction.

As in any valuation exercise, our final estimate of fiscal capacity depends on the cash flow projections, including the seigniorage revenue earned on Treasuries, and the discount rate assumptions. Both are subject to considerable uncertainty.

First, our measure of projected fiscal capacity relies on CBO projections of future primary surpluses as well as GDP and interest rate forecasts. The primary surplus projections are not traditional forecasts. To be concrete, Congress can pass new legislation in order to increase tax revenue or decrease non-interest spending. The CBO does not try to forecast such future fiscal policy adjustments. As we show in recent work (Jiang and others 2021), the CBO projections have systematically overstated realized surpluses over the past two decades. Should this overstatement continue, it would render our estimate of fiscal capacity overly generous. Even taking CBO projections at face value, our estimate of fiscal capacity suggests that large fiscal corrections relative to the CBO baseline are anticipated by US Treasury markets.[3] Alternatively, the market may be pricing in some form of real rate distortions or financial repression in the future.[4]

Second, our measurement of fiscal capacity relies on discount rates. We use the discount rates on a claim to GDP, or equivalently, the expected return on the unlevered market portfolio, to derive an upper bound on fiscal capacity. The estimate is sensitive to the discount rate. Choosing a lower discount rate results in higher estimates of fiscal capacity. To arrive at an estimate of fiscal capacity that matches the current valuation of the outstanding Treasury portfolio, we would need a discount rate that is lower than the projected growth rate of the economy. That would imply an implausibly low GDP risk premium and implausibly high valuations of other assets.[5] Lower discount rates also increase the sensitivity of fiscal capacity to interest

3. In a classic paper, Bohn (1998) argues that increases in the debt/output ratio predict larger future surpluses, but in a longer sample and after correcting for small-sample bias, Jiang and others (2021) find no evidence of this mechanism.

4. See Acalin and Ball (2022) for evidence on the role of real rate distortions through pegged nominal interest rates before 1951 in the postwar US fiscal experience. More recently, the Bank of Japan has been using yield curve control.

5. Put differently, we would need to engineer a violation of the TVC to match the valuation of Treasuries, given the CBO projections.

rate changes and worsen the duration mismatch. While the literature has argued that low interest rates increase fiscal capacity, the impact of low rates on duration mismatch has not received much attention.

Our forward-looking valuation approach, in the tradition of Hansen, Roberds, and Sargent (1991), is well-suited for use with the CBO projections. Others pursue a complementary backward-looking accounting approach to the question of fiscal sustainability which characterizes debt/output dynamics as a function of past returns and surpluses (Hall and Sargent 2011; Mehrotra and Sergeyev 2021). However, this approach is limited because it only considers the realized path of aggregate shocks.

Despite the secular decline in real rates, private investment has stagnated. This phenomenon has been referred to as the secular stagnation (Summers 2015). Economists have explored whether the US economy is dynamically inefficient, perhaps as a result of increased market power (Ball and Mankiw 2021; Aguiar, Amador, and Arellano 2021). Farhi and Gourio (2018) countered that risk premia may have increased as real rates have decreased, explaining the low private investment. When using deterministic models (without risk premia), economists may have mistakenly overestimated the net present value of private investment opportunities. Using stochastic models with substantial risk premia lowers the value of private investment opportunities. We make a related point about the government's fiscal capacity. In spite of the secular decline in real rates, the US government's fiscal capacity is more limited once risk premia are accounted for.

Government Ponzi schemes that look promising in deterministic economies typically do not survive exposure to aggregate risk and the presence of long-lived investors (Jiang and others 2020; Barro 2020). These schemes also do not survive a close look at the historical evidence which suggests that the fiscal capacity of governments has always been limited. For example, the United Kingdom, for which we have the longest continuous fiscal time series data, ran primary surpluses of 2.38 percent (1.22 percent) of GDP between 1729 and 1914 (1946). After 1946, the United Kingdom ran primary surpluses of 1.77 percent of GDP (Chen and others 2022). Our paper contributes to the measurement of these limits.

The paper is organized as follows. Section I describes the discounted cash flow analysis approach to measuring projected fiscal capacity and computes the latter in the benchmark scenario. Section II analyzes the effect of interest rate risk. Section III adds convenience yields. Section IV analyzes a front-loaded fiscal adjustment. Section V analyzes the hypothetical case of countercyclical tax revenue. The last section concludes.

I. Discounted Cash Flow Analysis

In a deterministic model without aggregate growth risk, the government can always roll over the debt when the risk-free rate is lower than the growth rate of the economy. The government's fiscal capacity may be unlimited.

This argument used in a deterministic setting does not carry over to an economy with priced aggregate growth risk for two reasons. First, the risk-free rate r_t^f cannot always be lower than the realized growth rate g_t. To see why, consider the case in which the aggregate growth rate is independently and identically distributed over time and the price/dividend ratio of a claim to GDP is constant. If r_t^f is always lower than g_t, then the return on going long in a claim to GDP (unlevered equity) and borrowing at the risk-free rate is always positive. Hence, we have created an arbitrage opportunity, not only for the government, but for all investors.[6]

Second, in a world with output growth risk, the Treasury portfolio is risk-free and earns the risk-free rate if and only if the tax claim is less risky than the spending claim (Jiang and others 2019). That restriction has teeth, and it appears to be violated in US data (Jiang and others 2020).[7] Our measure of fiscal capacity rules out free lunches for the government and investors. As pointed out by Lucas (2012), it is critically important to properly price risk when evaluating government policies.

In reality, going long in unlevered equity and short in the risk-free bond is quite risky. To be compensated for this risk, investors demand a large risk premium (Mehra and Prescott 1985). We call this the GDP or unlevered equity risk premium. This object plays a key role in our analysis.

In standard asset pricing models, the government debt is fully backed by future primary surpluses. The debt in 2021 is backed by primary surpluses $(\{T - G\}_{2022}^{2021+H})$, because the PDV of future debt, say H = two hundred

6. If $r_t^f < g_t$ in all states of the world, the return on a claim to output would always exceed the risk-free rate: $R_{t+1}^y = \dfrac{1 + pd}{pd}(1 + g_t) > 1 + r_t^f$, giving rise to unbounded profit opportunities for a long-lived investor borrowing at the risk-free rate and going long in unlevered equity. The scenario $r_t^f < g_t$ in all states of the world creates arbitrage opportunities not only for the government but for everyone else. One exception is the case of convenience yields λ_t, which drive Treasury yield below the true risk-free rate: $y_t = r_t^f - \lambda_t$. We discuss these in section III. Convenience yields decline when the debt/output ratio increases.

7. Moreover, the Treasury does not roll over the entire portfolio of debt every few months by issuing T-bills at the risk-free rate. The return on the portfolio of all outstanding Treasuries has exceeded the nominal growth rate of GDP throughout the 1980s and 1990s (Hall and Sargent 2011).

Figure 1. Government Balance Sheet: An Example

Panel A

	Assets	Liabilities
Until 2021 + H	$PV_{2021}(\{T\}_{2022}^{2021+H})$	$PV_{2021}(\{G\}_{2022}^{2021+H})$
After 2021 + H	$PV_{2021}(D_{2021+H}) \rightarrow \$0$	
		$D = PV_{2021}(\{T - G\}_{2022}^{2021+H})$

Panel B

	Assets	Liabilities
Until 2021 + H	$PV_{2021}(\{T\}_{2022}^{2021+H})$	$PV_{2021}(\{G\}_{2022}^{2021+H})$
After 2021 + H	$PV_{2021}(D_{2021+H}) \rightarrow \$0$	
		$D = PV_{2021}(\{T - G\}_{2022}^{2021+H} + D_{2021+H})$

Source: Authors' calculations.

years from now, in 2021 dollars, is arbitrarily small. This is often referred to as the no-bubble condition or the transversality condition (TVC).[8] Figure 1, panel A, illustrates the government's balance sheet in this standard setting.

Assumption 1: Debt is cointegrated with output. We assume that debt and output evolve together in the long run. Even when the current debt is risk-free (i.e., it has a beta of zero), future debt will be exposed to output risk because it is cointegrated with output. Hence, the discount rate applied to future debt, say in 2221, will include a GDP risk premium rp^y as well as a term premium ($r^f + term + rp^y$). When debt and output are cointegrated, the no-bubble condition is satisfied as long as the discount rate exceeds the growth rate ($r^f + term + rp^y - g > 0$), even when the risk-free rate is lower than the average growth rate ($r^f < g$). If we turned off all aggregate risk and set the risk premia to zero, the TVC condition would be violated when $r^f - g < 0$.

8. The TVC requires that the expected present-discounted value of debt in the far future, $\mathbb{E}_t[M_{t+T}D_{t+H}]$, goes to zero as the horizon H goes to infinity. The TVC is an optimality condition in an economy with long-lived investors. Jiang and others (2020) show that the TVC is satisfied as long as the GDP risk premium exceeds the gap between the growth rate and the risk-free rate.

If we push the horizon out far enough, under mild conditions, the current value of future debt goes to zero, $PV_{2021}(D_{2021+H}) \to 0$, and the value of debt equals the expected present-discounted value of future primary surpluses:[9]

$$D_{2021} = PV_{2021}\left(\left\{T - G\right\}_{2022}^{\infty}\right). \tag{1}$$

We define a country's projected fiscal capacity at the end of 2021 as the present-discounted value of future projected primary surpluses: $PV_{2021}(\{T - G\}_{2022}^{\infty})$.

This calculation requires an estimate of the future surpluses and an estimate of the discount rate. We tackle each of these in turn. We perform this calculation as of December 31, 2021. The actual market value of government debt at the end of 2021, D_{2021}, is 99.64 percent of GDP.

To be clear, there are models, typically without long-lived investors, that can generate bubbles in asset markets for long-lived assets, including bonds. In these models, there are no long-lived investors to enforce the TVC for long-lived assets.[10] Most of these models do not have priced aggregate risk.[11] In these models, the debt may not be fully backed by the PDV of surpluses. Instead, debt may be backed by future debt itself, as the PDV of future debt does not tend to zero. We can think of this as a rational bubble, as illustrated in figure 1, panel B.[12]

We analyze fiscal capacity while ruling out permanent bubbles in the Treasury market. First, many institutional investors with a long horizon such as endowments, pension funds, and sovereign wealth funds are active in US Treasury markets. Second, typically, these bubbles would also appear in other long-lived assets, such as stocks, resulting in implausible valuations for these assets. Third, nothing in these models singles out the United States as an ideal candidate for engineering these bubbles.

9. This equation is alternatively referred to as the government intertemporal budget constraint or the debt valuation equation. This equation has a long history, going back to seminal work by Hansen, Roberds, and Sargent (1991). This result, proven in Jiang and others (2019), follows from imposing (1) the government budget constraint in each period, (2) no-arbitrage conditions on individual bond prices, and (3) a transversality condition.

10. See Santos and Woodford (1997) for a classic reference.

11. See Dumas, Ehling, and Yang (2021) for a recent example.

12. Brunnermeier, Merkel, and Sannikov (2022) argue that the government can engineer violations of TVC by providing safe assets that serve uniquely as insurance against idiosyncratic risk.

I.A. Cash Flows

The cash flows we need are primary surpluses from 2022 onwards, that is, federal tax revenues minus federal non-interest spending. We break up this cash flow stream into the cash flow until 2052 and the cash flow after 2052. By value additivity, we can split up the PDV of surpluses as the sum of surpluses until the end of the CBO projection horizon in 2052 and the residual tail value:

$$PV_{2021}\left(\left\{S\right\}_{2022}^{\infty}\right) = PV_{2021}\left(\left\{S\right\}_{2022}^{2052}\right) + PV_{2021}\left(\left\{S\right\}_{2053}^{\infty}\right). \tag{2}$$

PRIMARY SURPLUSES UNTIL 2052 We use the CBO's long-term budget projections for the US federal government (CBO 2021a, Supplemental Table 1, Summary Data for the Extended Baseline). It contains the CBO projections for federal non-interest spending, revenues, debt held by the public, and GDP for each fiscal year from 2022 until 2051. These projections are as of May 2022. From the interest cost and debt projections, we can back out an implicit interest rate on the portfolio of outstanding government debt for those same years.

Table 1 lists the CBO's budget projections for the years 2022–2052 (CBO 2021a, 2021b). The first column reports government revenue as a percentage of GDP. The second column reports government spending excluding interest as a percentage of GDP. The third column reports the projected primary surplus as a percentage of GDP, given by column 1 minus column 2. The US federal government is projected to run large and growing primary deficits until the end of the projection window in 2052. Column 4 reports nominal GDP projections. For 2022 to 2032, we use projections from the May 2022 CBO report (CBO 2021c, 2022).[13] After that, we use the projected real GDP growth rate and the long-run projected rate of inflation.[14] We then compute the implied dollar numbers for projected nominal tax revenue and spending in columns 5 and 6. The CBO also projects interest costs and implied debt/GDP ratios for the federal debt held by the public. These are reported in column 10.[15]

13. The CBO provides a supplement to the May 2022 fiscal projection report called *An Update to the Budget and Economic Outlook: 2021 to 2031*.

14. Projections are from the figures in the CBO's May 2022 report *The 2022 Long-Term Budget Outlook*.

15. This excludes nonmarketable debt.

Table 1. Fiscal Capacity: Baseline Estimates

Year	T/Y (%) (1)	G/Y (%) (2)	$(T-G)/Y$ (%) (3)	Y (\$ billions) (4)	T (\$ billions) (5)	G (\$ billions) (6)	$y^{\$}$ (%) (7)	$r^{\$,y}$ (%) (8)	$PV(T-G)$ (\$ billions) (9)	D/Y (10)	D (\$ billions) (11)
2022	19.6	21.9	−2.3	24,694	4,836	5,405	0.42	3.02	(552.26)	97.9	24,173
2023	18.6	20.7	−2.0	26,240	4,889	5,419	0.76	3.36	(495.70)	96.0	25,193
2024	18.0	20.3	−2.2	27,291	4,924	5,535	0.99	3.59	(549.75)	96.1	26,217
2025	17.6	20.1	−2.5	28,271	4,982	5,696	1.15	3.75	(616.35)	97.5	27,561
2026	18.0	20.4	−2.3	29,266	5,280	5,962	1.27	3.87	(564.72)	98.8	28,925
2027	18.3	20.4	−2.2	30,332	5,548	6,201	1.36	3.96	(517.27)	100.0	30,326
2028	18.2	20.6	−2.4	31,487	5,716	6,486	1.43	4.03	(583.70)	102.0	32,105
2029	18.1	20.7	−2.6	32,716	5,934	6,773	1.49	4.09	(608.55)	103.2	33,760
2030	18.1	20.8	−2.7	33,996	6,161	7,066	1.55	4.15	(627.67)	105.3	35,808
2031	18.1	20.9	−2.7	35,318	6,402	7,371	1.59	4.19	(642.81)	107.5	37,949
2032	18.2	21.1	−2.9	36,680	6,662	7,722	1.63	4.23	(671.66)	109.6	40,213
2033	18.2	21.2	−3.0	38,081	6,938	8,062	1.67	4.27	(680.82)	112.0	42,636
2034	18.3	21.3	−3.0	39,519	7,217	8,413	1.71	4.31	(691.49)	114.4	45,219
2035	18.3	21.4	−3.1	40,996	7,506	8,779	1.74	4.34	(702.29)	117.0	47,975
2036	18.4	21.6	−3.2	42,514	7,801	9,166	1.77	4.37	(718.35)	119.8	50,926
2037	18.4	21.7	−3.3	44,074	8,110	9,567	1.80	4.40	(731.56)	122.7	54,088
2038	18.4	21.8	−3.4	45,680	8,423	9,975	1.83	4.43	(742.66)	125.8	57,472
2039	18.5	22.0	−3.5	47,335	8,749	10,391	1.85	4.45	(749.19)	129.1	61,087

2040	18.5	22.1	–3.6	49,035	9,082	10,827	1.88	4.48	(759.32)	132.5	64,963
2041	18.6	22.2	–3.6	50,782	9,426	11,272	1.90	4.50	(765.37)	136.1	69,115
2042	18.6	22.3	–3.7	52,581	9,782	11,727	1.92	4.52	(768.38)	139.9	73,568
2043	18.7	22.4	–3.8	54,443	10,158	12,208	1.94	4.54	(771.44)	143.9	78,343
2044	18.7	22.5	–3.8	56,372	10,539	12,685	1.96	4.56	(769.40)	148.0	83,447
2045	18.7	22.6	–3.9	58,371	10,939	13,193	1.98	4.58	(769.60)	152.3	88,909
2046	18.8	22.7	–3.9	60,444	11,359	13,709	2.00	4.60	(763.93)	156.7	94,734
2047	18.8	22.7	–3.9	62,594	11,798	14,219	2.01	4.61	(749.74)	161.2	100,911
2048	18.9	22.8	–3.9	64,824	12,260	14,782	2.03	4.63	(743.84)	165.8	107,481
2049	19.0	22.8	–3.9	67,132	12,726	15,328	2.04	4.64	(730.74)	170.5	114,436
2050	19.0	22.9	–3.9	69,514	13,217	15,900	2.05	4.65	(717.59)	175.2	121,798
2051	19.1	22.9	–3.8	71,970	13,733	16,500	2.07	4.67	(704.70)	180.1	129,588
2052	19.1	23.0	–3.9	74,505	14,254	17,130	2.07	4.67	(699.91)	185.0	137,852
Total PV									(21,160)		33,540

Source: Based on CBO projections released May 2022.

Note: Column 8 reports the discount rates used for spending and tax cash flows in that year. Column 9 reports an upper bound on the PDV of projected primary surpluses in 2021 $ billions. Column 10 reports the projected debt/GDP ratio for federal debt held by the public.

While the CBO forecasts GDP, inflation, and interest rates in unrestricted fashion, it makes projections of future revenues and non-interest spending based only on current law. The CBO assumes that temporary spending and tax changes will expire as provided in the law. However, the CBO projections assume that the federal government continues to pay for Social Security and Medicare even when the trust funds expire.[16]

Jiang and others (2021) document that CBO projections have been too optimistic over the past two decades. This was not true prior to the late 1990s. While some of the overly optimistic projections are no doubt due to the global financial crisis and the COVID-19 pandemic, the CBO projected a reduction in deficits well after the global financial crisis and before the COVID-19 pandemic that failed to materialize. If this pattern continues, our measure of projected fiscal capacity is likely to overstate the actual capacity.

We do not consolidate the Federal Reserve and the Treasury. Such a consolidation would not change the amount of government liabilities held by the private sector. It would merely imply a shortening of the maturity structure of the debt held by the private sector. Quantitative easing (QE) programs buy long-term Treasuries from the private sector and issue short-term bank reserves in return. The shorter maturity of the debt held by the public would further exacerbate the maturity mismatch we highlight below. The consolidation would not affect the PDV of projected future surpluses.

I.B. Discount Rates

Our approach confronts risk head-on by using discount rates that reflect the cash flow risk in future spending, tax revenue, and future debt outstanding.

RISKINESS OF TAX REVENUES AND NON-INTEREST SPENDING The CBO projections for future non-interest spending and tax revenue in table 1 are point estimates; there is substantial uncertainty around the point estimates. This uncertainty is naturally related to the uncertainty in the underlying macroeconomy. Because the underlying cash flows are risky, they cannot be discounted off the Treasury yield curve. As in any valuation exercise, the proper

16. The non-payable part of Social Security and Medicare remain liabilities for the government even after the corresponding trust funds are exhausted. We are grateful to Phillip Swagel and Molly Dahl for explaining the CBO's approach: "The CBO's extended baseline projections follow the agency's ten-year baseline budget projections and then extend most of the concepts underlying those projections for an additional twenty years. In accordance with statutory requirements, the CBO's projections reflect the assumptions that current laws generally remain unchanged, that some mandatory programs are extended after their authorizations lapse, and that spending on Medicare and Social Security continues as scheduled even if their trust funds are exhausted."

discount rate needs to reflect the systematic riskiness of the cash flows. The key question then becomes: What is the underlying source of aggregate risk to primary surpluses?

To develop some intuition, consider the simplest case in which government spending and tax revenue are a constant fraction of GDP. Then, by definition, these claims are exactly as risky as a claim to GDP. The latter is often referred to as the total wealth or market portfolio (Jensen 1972; Roll 1977; Stambaugh 1982; Lustig, Van Nieuwerburgh, and Verdelhan 2013). The return on the total wealth portfolio plays a central role in the canonical capital asset pricing models (CAPM), ranging from the Sharpe-Lintner CAPM to the version of the Breeden-Lucas-Rubenstein consumption-based CAPM with long-run risks developed by Bansal and Yaron (2004). The total wealth return is often proxied in the asset pricing literature by the unlevered return on the stock market. The idea is that a portfolio that invests in all publicly listed companies broadly reflects the evolution of the overall economy.[17] We will adopt this approach, recognizing that the stock market is a levered claim to corporate cash flows. This will lead us to un-lever the equity return to arrive at the total wealth return, the return on a claim to future GDP. We discuss the implementation below.

Modeling tax revenue and non-interest spending as a constant fraction of GDP is sensible in the long run. At business cycle frequencies, the ratio of tax revenue to GDP is pro-cyclical while the ratio of non-interest spending to GDP is countercyclical (Jiang and others 2019). These cyclical patterns imply that a claim to all future tax revenues is riskier than a claim to all future GDP, while a claim to all future non-interest spending is safer than the GDP claim. Intuitively, the spending claim is a hedge that has high payoffs in bad states of the world (recessions, high stochastic discount factor, M, states). Investors prefer such hedges, bidding up their price, and bidding down their expected return. The tax revenue claim has the opposite properties, where tax revenues rise as a share of GDP exactly when investors care least about the extra income (good times, low M states). Hence the tax claim is riskier than a claim to GDP, just like the dividend claim on stocks is riskier than the GDP claim. It carries an expected return and risk premium that exceeds that on the GDP claim. In summary, in the short run, the tax (spending) claim is exposed to more (less) business cycle risk.

17. This effectively assumes that the aggregate dividends from all publicly listed firms have the same riskiness as all corporate cash flows. Publicly traded firms represent a sizeable share of aggregate corporate cash flows. If anything, shares in the private firms have higher expected returns, because of the illiquidity. As a result, our approach provides a lower bound on the market risk premium.

In the long run, spending and taxes are both co-integrated with output, and hence (equally) exposed to long-run output risk.[18]

Assumption 2: Government taxes, spending, and the value of debt are co-integrated with output. Co-integration is a necessary condition for fiscal sustainability. When fiscal policy is sustainable, then taxes, spending, debt, and output are co-integrated with output. As a result, surpluses are more risky than output in the short run and equally risky in the long run.

Combining the short-run and long-run risk properties, we find that the tax claim is riskier than the GDP claim, which is riskier than the spending claim.[19]

This gives us the following result: the true discount rate for projected tax cash flows is higher than the discount rate for projected spending cash flows: $\mathbb{E}[r^T] \geq \mathbb{E}[r^y] \geq \mathbb{E}[r^G]$, because tax revenue (spending) is riskier (safer) than GDP.

Importantly, this result immediately implies that the government debt portfolio cannot have a zero beta, that is, it cannot be risk-free. The debt will have a positive beta, that is, it will carry a positive risk premium.

UPPER BOUND ON FISCAL CAPACITY Our approach is to compute an upper bound on fiscal capacity. This upper bound obtains when discounting future non-interest spending and tax revenue at the same discount rate, namely, the expected return on a claim to GDP: $\mathbb{E}[r^T] = \mathbb{E}[r^G] = \mathbb{E}[r^y]$.

Assumption 3: To derive an upper bound, we assume that future spending, tax revenue are all as risky as GDP. We use the following discount rates: $\mathbb{E}[r^T] = \mathbb{E}[r^y] = \mathbb{E}[r^G]$.

By using the same discount rate for the tax and spending claims, we maximize the value of the tax claim because we use a discount rate that is too low, and we minimize the value of the non-interest spending claim because we use a discount rate that is too high. Overstating the value of the tax claim and understating the value of the non-interest spending claim results in a value of the primary surplus claim that is unambiguously too large,

18. A strip is a claim to one dividend payment in the future. When taxes (spending) are co-integrated with GDP, then long-run returns on tax strips and output strips converge; see proposition 3 in Jiang and others (2019). See Backus, Boyarchenko, and Chernov (2018) for a general proof. In the long run, the tax claim, the spending claim, and the output claim are all equally risky.

19. As explained by Jiang and others (2019), this rules out that the entire debt portfolio has zero or negative beta. Generating zero-beta debt can be achieved only if the beta of the tax claim is lower than the beta of the spending claim, that is, by rendering the tax claim less risky than the spending claim. The empirical evidence points in the opposite direction. In addition, highly persistent deficits are inconsistent with risk-free debt when the debt/output policy is mean-reverting. See also van Wijnbergen, Olijslagers, and de Vette (2021) and Barro (2020).

thus deriving an upper bound on the fiscal capacity.[20] In other words, our measure will tend to overstate fiscal capacity.

IMPLEMENTATION: MEASURING THE GDP RISK PREMIUM As argued above, we proxy a claim to GDP as the unlevered version of a claim to the dividends of all publicly listed stocks. Hence, to construct $\mathbb{E}[r^y]$, we begin by constructing a measure of the expected return on equity and un-lever this expected return in a second step.

We infer the expected return on a claim to equity from valuations in the stock market. There are many ways one could measure the expected return on stocks: from a vector autoregressive model, as in Jiang and others (2019); from survey expectations (Fernandez, Bañuls, and Acin 2021); or from option markets (Andersen, Fusari, and Todorov 2015; van Binsbergen, Brandt, and Koijen 2012), to name a few.

For simplicity, we use an off-the-shelf estimate from the private sector. It is an average of two approaches to measure the expected real return on US equities going forward, as of the end of 2021: an earnings-based and a payout yield–based estimate.[21] The earnings-based estimate for the expected real return on US stocks is given by the payout ratio times the earnings/price ratio plus the projected growth rate of earnings:

$$(3) \qquad \mathbb{E}\left[r^{equity}\right] = D/E \times E/P + g_{EPS} = 0.5 \times 2.8\% + 1.5\% = 2.9\%,$$

where we use the inverse of Shiller's CAPE ratio to measure the earnings/price ratio, a dividend payout ratio of 0.5, and an expected growth rate in earnings per share of 1.5 percentage points, all measured at the end of 2021. The payout yield–based estimate for the real expected return on US stocks is given by:

$$(4) \quad \mathbb{E}\left[r^{equity}\right] = D/P + NBY + g_{PAGG} = 1.3\% + 0.2\% + 2.7\% = 4.2\%,$$

where D/P is the dividend yield on the S&P 500, NBY is the net buyback yield, and g_{PAGG} is a forecast of aggregate US earnings growth, also measured at the end of 2021. We combine these two estimates with equal weights to obtain a blended real expected return of 3.6 percent. The real risk-free

20. Our approach is to estimate the expected return on the tax claim and the spending claim by committing to a fully specified asset pricing model as well as dynamics for fiscal cash flows. This is the first approach pursued by Jiang and others (2019).

21. The approach was developed by AQR for its capital market assumptions; see Portfolio Solutions Group (2022), for details.

return is estimated to be −1.5 percent. As a result, we obtain an estimate of 5.1 percent in excess of the risk-free rate. This number is very close to the 5.5 percent average (and median) estimate of the US equity risk premium from a recent academic survey (Fernandez, Bañuls, and Acin 2021).

The equity risk premium is the risk premium on a levered claim. We are interested in the risk premium on an unlevered claim. The debt/equity ratio for the US non-financial corporate sector is roughly 1:2 at the end of 2021, so that the equity/asset ratio is 2:3. As a result, we obtain an unlevered equity premium of 3.4 percent from a levered equity premium of 5.1 percent (two-thirds of 5.1 percent is 3.4 percent). This assumes a zero risk premium on corporate debt.

We also compute an expected excess return of long-term bonds of 0.8 percent. This means that unlevered equities earn a risk premium rp^y of 2.6 percent over long-term bonds. This is our measure of the GDP risk premium. The 2.6 percent GDP risk premium we use here is close to the 2.9 percent GDP risk premium that comes out of the calibrated disaster model in Jiang and others (2020). It is also close to the 2.4 percent risk premium on the total wealth claim obtained by Lustig, Van Nieuwerburgh, and Verdelhan (2013).[22]

We argue that 2.6 percent is a low estimate of the annual GDP risk premium for two reasons. First, the average excess return on stocks has been 8 percent over the 1947–2021 period and may have been at a cyclical low at the end of 2021.[23] Hence the unlevered equity risk premium was unusually low at the time of our measurement. Second, using a higher cost of debt for corporations than the risk-free rate (assuming a positive corporate bond risk premium when un-levering) would also increase the unlevered equity risk premium. Using a lower discount rate will increase our measure of fiscal capacity. This will result in a conservative estimate of projected fiscal capacity, given that we will show that even this generous estimate of fiscal capacity falls short of the outstanding amount of debt at the end of 2021.

To construct the discount rates for discounting tax revenue and spending claims at each horizon h, we start from the nominal zero-coupon bond yield curve at the end of 2021 for maturities from one to thirty years, constructed

22. The latter estimate recognizes that a claim to GDP is potentially different from a claim to the cash flows of all current businesses, because the businesses in the current cohort are short-lived.

23. Sample averages calculated with data from the Center for Research in Security Prices, LLC, “Data Access Tools,” https://www.crsp.org/products/software-access-tools.

and updated by Gürkaynak, Sack, and Wright (2007), and then add the output risk premium of 2.6 percent:

$$\mathbb{E}\left[r^{\$,y}\left(h\right)\right] = y_t^{\$,f}\left(h\right) + rp^y. \tag{5}$$

This discount rate is reported in column 8 of table 1, with the zero-coupon nominal bond yield component of that discount rate listed in column 7.[24]

I.C. Steady-State Fiscal Capacity

As a warm-up exercise, we compute a measure of steady-state fiscal capacity. In the steady state, the government runs a constant primary surplus relative to GDP. Given that the tax claim is riskier than the spending claim, an upper bound on the steady-state fiscal capacity is given by the valuation ratio on a claim to GDP times the steady-state surplus. In the steady state, the valuation of future surpluses is given by the price/dividend ratio on a claim to GDP times the steady-state surplus:

$$PV_{2021}^{upper,ss}\left(\left\{T - G\right\}\right) = \frac{S}{Y}\sum_{j=1}^{\infty}\frac{\mathbb{E}_{2021}\left(Y_{2021+j}\right)}{\left(1 + r^{\$,y}\right)^j} = pd^y \times \frac{S}{Y} \times Y_{2021}. \tag{6}$$

We use the thirty-year zero-coupon yield at the end of 2021 to proxy for the long end of the Treasury yield curve, and we use the CBO's long-run forecast for real growth of 1.5 percent and inflation of 2 percent. The nominal long discount rate minus the nominal growth rate is given by:

$$\begin{aligned} r^{\$,y} - g &= y_{2022}^{\$,f}\left(30\right) + rp^y - g \\ &= 2.07\% + 2.60\% - \left(1.50\% + 2\%\right) = 1.17\%. \end{aligned} \tag{7}$$

We can use Gordon's growth formula to compute the valuation ratio for the claim to GDP:

$$pd^y = \frac{1}{r^{\$,y} - g} = \frac{1}{1.17\%} = 85.8. \tag{8}$$

24. We assume that the yield on a thirty-one-year zero-coupon bond equals the yield on a thirty-year bond.

Table 2. US Treasury Balance Sheet in Steady-State Example

Assets		*Liabilities*	
$PV_{2021}(\{T\})/Y_{2021}$	$19.78 = 23.06\% \times 85.8$	$PV_{2021}(\{G\})/Y_{2021}$	$18.79 = 21.9\% \times 85.8$
		D/Y_{2021}	$0.99 = 1.16\% \times 85.8$
Total	19.78	Total	19.78

Source: Authors' calculations.

Note: Market values are expressed as a multiple of US GDP at the end of 2021. The steady-state example is based on the actual spending/GDP ratio in 2022.

The multiple on a claim to GDP is 85.8.[25] An unlevered company whose cash flows grow at the same rate as the US economy would have a price/dividend ratio of 85.8 in 2021.[26] At this high multiple, total US wealth is about 85.8 times the size of GDP.[27] This historically high multiple reflects low rates and low risk premia at the end of 2021.

Table 2 shows the US Treasury's balance sheet in market values, expressed as a percentage of GDP. Total assets and total liabilities are exposed to the same cash flow risk. The Treasury cannot financially engineer risk away. The risk in the tax process on the left-hand side of the ledger has to show up on the right-hand side in spending risk or in the riskiness of the debt. If the primary surplus/GDP ratio, S/Y, is constant, then the surplus inherits the risk properties of a GDP claim. In this simple case, the discount rate for a GDP claim is the right discount rate for the surplus claim. And the valuation of debt would be 0.99 times GDP, as shown in table 2. However, as we have explained, S/Y is actually pro-cyclical in the data, implying that the surplus claim is riskier than the output claim. As a result, our calculation produces an upper bound on fiscal capacity.

Under "Liabilities" in table 2, we start from the 2022 spending ratio of 21.9 percent. We need a steady-state primary surplus of 1.16 percent of GDP to get to an upper bound on fiscal capacity that includes the observed debt/GDP ratio of 99.7 percent as of the end of 2021: 85.8×1.16 percent = 99.7 percent of 2021 GDP. Under "Assets," we back out the implied steady-state tax ratio T/Y of 23.06 percent that is needed. The implied value of the tax claim is almost twenty times GDP.

25. Using a different approach with a no-arbitrage term structure model, Lustig, Van Nieuwerburgh, and Verdelhan (2013) obtain an average US wealth/consumption ratio of 83, a similar value.

26. If that company were only expected to exist for fifty years, the multiple would still be 64.5.

27. In 2021, that's about \$5.8 million per American. Most of this is the PDV of future labor income.

The United States gets an additional 85.8 percent of GDP in fiscal capacity (maximum) per 1 percent of steady-state primary surplus S/Y. As noted above, our GDP risk premium estimate is low, resulting in a high price/dividend ratio on the GDP claim. As a result, our calculation produces high estimates of fiscal capacity, holding fixed the projected surpluses. In addition, the secular decline in real rates over the past decades has boosted US fiscal capacity per percentage point of primary surplus.

However, the CBO does not project any surpluses over its projection horizon. Column 3 in table 1 reports the actual projected primary deficits. The CBO projects an average deficit of 3.19 percent of US GDP between 2022 and 2052. One would need a large, permanent fiscal correction of 4.35 percent of GDP (from −3.19 percent to 1.16 percent) to reconcile this back-of-the-envelope upper bound with the actual value of US Treasury debt/GDP. For this to work out exactly, the steady-state surplus/GDP ratio would have to be acyclical.

I.D. Baseline Estimate of Fiscal Capacity

Next, we carry out our main analysis, which is to compute fiscal capacity as spelled out in equation (11). We discount each CBO projected cash flow, column 5 minus column 6 of table 1, with the discount rate $r^{\$,y}(h)$, shown in column 8, to arrive at the present discounted value listed in column 9.[28] The sum of the PDV of primary surpluses from 2022–2052 adds up to −\$21.16 trillion:

$$(9) \qquad PV_{2021}^{upper}\left(\left\{T - G\right\}_{2022}^{2052}\right) = \sum_{h=1}^{31} \frac{T_{2021+h} - G_{2021+h}}{\left(1 + r^{\$,y}\left(h\right)\right)^{h}} = -\$21.16 \text{ trillion.}$$

This is the sum of column 9 starting with −\$552 billion, the PDV of deficit in 2022, until and including −\$699.9 billion, the PDV of the 2052 deficit.

According to the CBO debt projections, reported in column 10 of table 1, the debt outstanding will equal 185 percent of US GDP at the end of 2052. This would amount to approximately \$138 trillion in nominal debt, as shown in column 11.

We assume that surpluses are a constant fraction of GDP in each year after 2053. Furthermore, we impose that equation (1) holds at the end of

28. Alternatively, we could discount projected cash flows in constant dollars using the yields on real zero-coupon bonds. The results are quite similar.

2052, namely, that the projected debt/output ratio in 2052 (see column 10) is fully backed by surpluses:

$$\left(\frac{D}{Y}\right)_{2052} = \frac{S}{Y} \times PV_{2052}\left(\left\{Y\right\}_{2053}^{\infty}\right). \tag{10}$$

Given that we have the CBO's projection for the debt/GDP ratio at the end of 2052, we can back out what constant surplus/GDP ratio is needed in the years after 2052 to satisfy equation (10). This implied surplus/GDP ratio will be positive since the projected debt/GDP ratio in 2052 is 185 percent of GDP, as shown in the last row of column 10 of table 1.

What do we need to assume about surpluses starting in 2053 to justify this number as the present-discounted value of future primary surpluses, as in equation (10)? Recall that the multiple on a claim to GDP at the end of 2021 is 85.8. It seems reasonable and conservative to use this same multiple at the end of 2052. The valuation multiple of 85.8 at the end of 2021 is high relative to its historical mean because of low long-term nominal rates and a low risk premium, and it is likely to revert back to its long-run mean. Using the historical average multiple would result in a higher required annual average primary surplus after 2052 to justify the same debt/output ratio at the end of 2052. This does not affect the present value of debt in 2052, only the required surpluses to repay this debt. To obtain a valuation of the debt outstanding at the end of 2052 equal to 185 percent of GDP, the US federal government would need to generate an annual primary surplus of 2.16 percent after 2052 (2.16 percent × 85.8 = 185 percent).

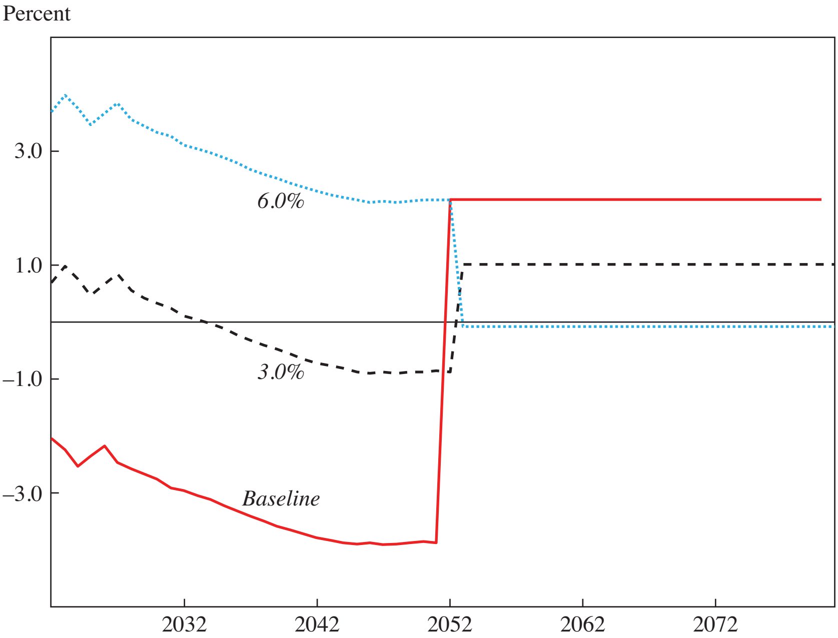
Assuming equation (10), our 2021 fiscal capacity estimate in equation (2) can be rewritten as the sum of the PDV of primary surpluses until the end of the projection horizon and the PDV of outstanding (projected) debt:

$$PV_{2021}\left(\left\{S\right\}_{2022}^{\infty}\right) = PV_{2021}\left(\left\{S\right\}_{2022}^{2052}\right) + PV_{2021}\left(D_{2052}\right). \tag{11}$$

Figure 2 plots the time path of projected primary surpluses; until 2052, it plots the projected primary surpluses from the CBO. After 2052, the primary surplus is assumed to be equal to 2.16 percent, the surplus needed to enforce the intertemporal budget constraint at the end of 2052.

The debt outstanding at the end of 2052, projected to be 185 percent of GDP, also needs to be discounted back to 2021 using the same discount rate

Figure 2. CBO Projections of Primary Surplus



Sources: CBO and authors' calculations.

Note: Shown are baseline CBO projections of primary surplus for 2022–2052, followed by primary surpluses after 2052 needed to pay back the debt in 2052; primary surpluses between 2022–2052 increased by 3 percent of GDP each year, followed by primary surpluses after 2052 needed to pay back the debt in 2052; and primary surpluses between 2022–2052 increased by 6 percent of GDP each year, followed by primary surpluses after 2052 needed to pay back the debt in 2052.

used for the primary surplus cash flow in 2052. The second term in equation (11) is given by:

$$PV_{2021}^{upper}\left(D_{2052}\right) = \left(\frac{D}{Y}\right)_{2052} \times \frac{\mathbb{E}_{2021}\left(Y_{2052}\right)}{\left(1 + r_{31}^{\$,y}\right)^{2052-2021}} = \$33.54 \text{ trillion.} \quad (12)$$

In our approach, future debt is assumed to be as risky as output. The future debt/output ratio in 2052 is constant across all possible output growth paths, but, of course, the debt itself is subject to GDP growth risk. The discount rate we use is the one appropriate for the stochastically growing GDP claim. Discounted back to the end of 2021, the PDV of D_{2052} is \$33.5 trillion.

When we add up the discounted value of debt outstanding in 2052 and the surpluses between 2022 and 2052, we obtain our baseline fiscal capacity estimate of \$12.38 trillion:

$$PV_{2021}^{upper}\left(\left\{T - G\right\}_{2022}^{2052}\right) + PV_{2021}^{upper}\left(D_{2052}\right) = -\$21.16 + \$33.54 \tag{13}$$

$$= \$12.38 \text{ trillion.}$$

The key observation is that this fiscal capacity estimate falls about \$10 trillion short of the actual valuation of debt in 2021 of \$22.3 trillion. In sum, our projected fiscal capacity bound cannot be reconciled with the actual valuation of debt at the end of 2021, given the baseline CBO projections of future primary surpluses, debt, and realistic discount rates.

This is a surprising result in light of four observations that bear repeating. First, this is an upper bound on fiscal capacity by virtue of discounting the fiscal cash flows at the GDP discount rate (rather than at the higher tax and lower spending discount rates). Second, the CBO's primary surplus projections have tended to be too high compared to realized values over the past two decades. Third, our point estimate for the GDP risk premium is, if anything, low. Fourth, we have assumed that the United States will generate primary surpluses after 2052 that are large enough to rationalize the projected value of outstanding debt in 2052. This would constitute a sea change from what we have observed in the past many decades. Relaxing any of these four assumptions would result in an even lower value for projected fiscal capacity and an ever larger wedge between the estimated fiscal capacity and the observed debt/GDP ratio at the end of 2021. Those are the four reasons that our estimate of projected fiscal capacity is conservative—if anything, too high rather than too low.

I.E. Discounting Future Debt

The right discount rate for debt outstanding far in the future includes the GDP risk premium when output and debt are co-integrated. The reason is that GDP in the far future is uncertain, and hence risky.[29] In our calculation, we use 2.07 percent for the nominal long yield, 2.60 percent for the GDP risk premium, and 3.50 percent for the long-run nominal growth rate g. These values imply that the TVC is satisfied (4.67 percent > 3.5 percent).

29. If the debt/output ratio is stationary, the necessary condition for TVC to be satisfied, $lim_{H\to\infty})\mathbb{E}_t[M_{t+h}D_{t+h}] = 0$, is $y^{\$,f}(H) + rp^y > g + \frac{1}{2}\sigma^2$ for some long horizon H and where σ is the volatility of output growth; see Jiang and others (2020), for details.

Importantly, this long-run discount rate that includes a GDP risk premium is the right discount rate for future debt regardless of the short-term debt/output, tax, and spending dynamics, and even when the current debt is risk-free, that is, has a zero beta.

If we had used the risk-free yield curve without adding the GDP risk premium when discounting future debt, then the discounted value of future debt in 2052 would have been \$73.15 trillion in 2021 dollars. The present value of the deficits until 2052 would have been –\$33.15 trillion. We would have obtained a fiscal capacity estimate of \$40 trillion at the end of 2021, comfortably above the observed debt/GDP ratio at the end of 2021. The federal government's debt is projected to grow faster than output, and the discount rate (2.07 percent) is lower than the growth rate of output (3.50 percent). This is essentially the $r < g$ approach to fiscal sustainability. As we push the final period T farther out, the PDV of debt outstanding at T does not converge to zero.

From a standard finance perspective, the $r < g$ argument is flawed, unless the GDP risk premium is zero. Future debt outstanding cannot be discounted using the risk-free yield curve unless the future debt's valuation is known today or unless its valuation is insensitive to the growth rate of output. This cannot be the case when debt and output are co-integrated, a necessary condition for fiscal sustainability (see assumption 2), even if current debt is risk-free (zero beta). As a result, discounting future debt at the risk-free rate is not consistent with fiscal sustainability. When discounted at a discount rate that includes the GDP risk premium, the value of future debt is much smaller, and the fiscal capacity estimate does not increase if we push T out farther into the future.

Suppose we took the counterfactual view that the entire debt portfolio really had a zero beta, because the tax claim was less risky than the spending claim. Then we could discount the projected surpluses until 2052 off the risk-free yield curve. However, we would still need to discount the future debt at the proper discount rate, which includes the GDP risk premium. The estimated projected fiscal capacity would then become:

$$(14) \qquad PV_{2021}^{upper}\left(D_{2052}\right) = \frac{D_{2052}}{\left(1 + r_{31}^{\$,y}\right)^{2052-2021}} = -\$33.74 + \$33.54\,t$$
$$= -\$0.20 \text{ trillion.}$$

We would end up at near-zero fiscal capacity, because the projected deficits increase in present value when discounted at a lower rate. This calculation shows that even discounting future primary surpluses over the next thirty years at the risk-free rate results in a low estimate of fiscal capacity as long as debt in the far future is discounted using a conceptually coherent discount rate.

This discussion raises a related question: How low would the GDP risk premium have to be to result in a fiscal capacity estimate that matches the observed debt/GDP ratio at the end of 2021? The answer is 1.37 percent per year. However, at this risk premium, the TVC fails because the discount rate is lower than the GDP growth rate and the economy is dynamically inefficient:

$$(15)\quad r^{\$,y} - g = y^{\$,f}_{2022}(30) + rp^{y} - g = 2.07\% + 1.37\% - (1.50\% + 2\%) < 0.$$

The steady-state multiple on claim to GDP tends to ∞. This has troubling valuation implications. An unlevered firm whose cash flows are expected to grow at the rate of US output growth would have an infinite valuation. We conclude that a value of 1.37 percent or lower for the GDP risk premium is implausibly low. In the baseline scenario, we cannot match the valuation of debt without engineering a violation of the TVC.

I.F. Reverse Engineering

Given our assumptions and and the result noted under assumption 2, the debt cannot be risk-free. The CBO assumes that the debt can be rolled over until 2052 at the projected interest rates. Even though the CBO does project an increase in interest rates in the long term, its projected interest rates may not be consistent with the true risk characteristics of the debt, implied by our analysis. The calculation of our benchmark fiscal capacity measure above, which takes the CBO interest rate projections until 2052 as given, can then be interpreted as consistent with the notion of persistent mispricing.

Alternatively, we can insist that the debt be priced correctly today given the CBO projections. Instead of using the CBO's projected debt/output ratio, we can back out the steady-state surplus after 2052 that is needed in order to obtain an estimate for fiscal capacity at the end of 2021 that equals the market value of outstanding debt:

$$(16)\qquad PV^{upper}_{2021}\left(\{T - G\}^{2052}_{2022}\right) + PV^{upper}_{2021}\left(D_{2052}\right) = -\$21.16 + \$43.45$$
$$= \$22.29 \text{ trillion.}$$

To obtain \$43.45 trillion for the present value of debt in 2052, we need annual primary surpluses of 2.79 percent from 2053 onward:

$$PV_{2021}^{upper}\left(\left\{T-G\right\}_{2022}^{2052}\right)/Y_{2052}=\frac{S}{Y}\times PV_{2052}\left(\left\{Y\right\}_{2053}^{\infty}\right) \tag{17}$$

$$= 2.79\% \times 85.8 = 239\%.$$

This can be interpreted as a debt/output ratio in 2052 of 239 percent, instead of the 185 percent projected by the CBO.

What explains the difference with the CBO projection of 185 percent? If we roll over the debt at the GDP discount rate in column 8 of table 1 until 2052, instead of using the CBO projected interest rates, the projected debt/output ratio is 239 percent rather than 185 percent. This reverse engineering exercise imposes that the debt be correctly priced and that the interest rates the Treasury pays on the debt reflect the risk.

II. Interest Rate Risk

II.A. Duration

The duration of the primary surplus claim is very high in the baseline scenario because the surpluses are extremely back-loaded; recall the baseline in figure 2. The Macaulay duration of the surplus claim is 283.2 years.[30] Figure 3 plots the contribution of each payment at horizon k to the total duration $\frac{k \times PV\left(S_{2021+k}\right)}{\sum_{h=1} PV\left(S_{2021+h}\right)}$. The duration is the sum of all bars.

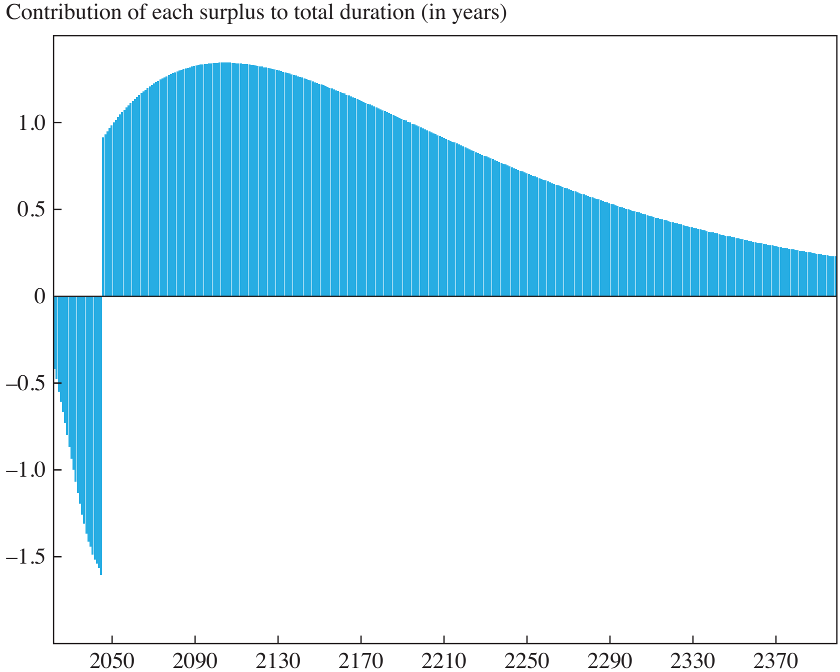
Given this high duration of the surplus claim, US fiscal capacity is very sensitive to the yield curve. We present two sets of calculations, one for a hypothetical 100 basis point parallel shift up in the yield curve, and one for the actually observed changes in the yield curve in the first five months of 2022.

II.B. Parallel Shift in Yield Curve

We study a 100 basis point parallel upward shift in the yield curve, holding constant all other parameters, including nominal GDP growth. Increasing

30. If we (somewhat implausibly) assume that the Treasury pays back all outstanding debt at the end of 2052 in one large payment rather than with gradual future surpluses, then the duration of the surplus claim becomes 44.7 years.

Figure 3. Duration Composition in Baseline Scenario



Source: Authors' calculations.

Note: Contribution of each payment $\frac{K \times PV(S_{2021+k})}{\sum_{h=1} PV(S_{2021+h})}$ to the total duration in the CBO baseline projection. The duration (measured in years) is the sum of the plotted contributions.

interest rates while holding nominal GDP growth constant amounts to an increase in the real growth-adjusted yield, that is, in $r - g$. This increase could reflect, for example, the unwinding of QE programs.[31] The upward shift in yields increases the discount rate of future surpluses and of future debt by 100 basis points, as shown in columns 8 and 9 of table 3. We also add an additional 100 basis points to the CBO's projected net interest cost as a fraction of debt in each year between 2022 and 2052, as shown in

31. Economists have found that large-scale asset purchases by the Federal Reserve have successfully lowered long-term yields (Krishnamurthy and Vissing-Jorgensen 2011; D'Amico and others 2012; Joyce and others 2011), with estimates ranging from 50–100 basis point declines. This implies that in the absence of QE, nominal long-term bond yields would be higher by that amount. The assumption that the GDP risk premium does not change is consistent with a narrow convenience yield view, as discussed in section III.

column 4 of table 3.[32] This extra interest cost affects the debt dynamics via $D_{t+1} = D_t \times R_{t+1} + S_{t+1}$. We compute these projected debt dynamics using the original projected primary surpluses and the CBO's interest rate projections plus 100 basis points.

The projected debt outstanding in this high-rate scenario grows to 223 percent of GDP in 2052 or to \$166.17 trillion, as shown in columns 11 and 12. Because of the 100 basis point rate increase, the steady-state multiple of a claim to GDP decreases from 85.8 to 46.2. Starting in 2053, the United States now has to generate a steady-state primary surplus of 4.83 percent (= 223 percent ÷ 46.2), an increase of 2.67 percentage points of GDP relative to the corresponding number in the baseline scenario of 2.16 percentage points of GDP, that is, before the interest rate change.[33] Hence, an increase in rates of 100 basis points, holding constant nominal GDP growth, implies an increase of 2.67 percentage points of GDP in annual surpluses starting in 2053. The increase in surpluses starting in 2053 divided by the increase in rates is 2.67. This multiple is the signature of the duration mismatch on the Treasury's balance sheet.

A dramatic increase in long-run future surpluses is one adjustment mechanism in response to the interest rate increase. Alternatively, if investors believe the government is unable to generate surpluses of this size, the valuation of the Treasury portfolio has to decline, triggering a sell-off and a widening of default spreads.

As mentioned, one can reverse-engineer the GDP risk premium that sets the fiscal capacity equal to the market value of debt. If we had assumed—counterfactually—that the GDP risk premium was 1.37 percent per year rather than 2.60 percent per year, the duration of the surplus claim would be 651, more than twice the baseline value. While a lower GDP risk premium increases fiscal capacity, it increases the sensitivity of that fiscal capacity to increases in interest rates. From a policy perspective, this means that duration and rollover risk are especially high when discount rates are low.

II.C. Higher Interest Rates in 2022

The first several months of 2022 saw a dramatic increase in interest rates. Between December 31, 2021, and May 31, 2022, the two-year zero-coupon

32. The CBO reports net interest/GDP and GDP projections from which we back out an estimate of the effective interest rate on debt R_t.

33. The estimate of the upper bound on fiscal capacity is now at \$12.03 trillion, which is close to the baseline number. The key point, however, is that this assumes 4.83 percent of GDP in primary surplus starting in 2053 compared to 2.16 percent of GDP in the baseline case.

Table 3. Fiscal Capacity with Higher Interest Rates

Year	*T/Y* *(%)* *(1)*	*G/Y* *(%)* *(2)*	*(T – G)/Y* *(%)* *(3)*	*NI/D* *(%)* *(4)*	*Y* *($ billions)* *(5)*	*T* *($ billions)* *(6)*	*G* *($ billions)* *(7)*	$y^{\$}$ *(%)* *(8)*	$r^{\$,y}$ *(%)* *(9)*	*PV(T – G)* *($ billions)* *(10)*	*D/Y* *(11)*	*D* *($ billions)* *(12)*
2022	19.6	21.9	–2.3	2.8	24,694	4,836	5,405	1.42	4.02	(546.95)	95.1	23,475
2023	18.6	20.7	–2.0	2.8	26,240	4,889	5,419	1.76	4.36	(486.24)	94.0	24,669
2024	18.0	20.3	–2.2	3.1	27,291	4,924	5,535	1.99	4.59	(534.13)	95.4	26,040
2025	17.6	20.1	–2.5	3.3	28,271	4,982	5,696	2.15	4.75	(593.15)	97.7	27,615
2026	18.0	20.4	–2.3	3.5	29,266	5,280	5,962	2.27	4.87	(538.31)	100.0	29,256
2027	18.3	20.4	–2.2	3.6	30,332	5,548	6,201	2.36	4.96	(488.39)	102.1	30,967
2028	18.2	20.6	–2.4	3.8	31,487	5,716	6,486	2.43	5.03	(545.90)	104.5	32,907
2029	18.1	20.7	–2.6	3.9	32,716	5,934	6,773	2.49	5.09	(563.74)	107.0	35,022
2030	18.1	20.8	–2.7	4.0	33,996	6,161	7,066	2.55	5.15	(575.94)	109.8	37,322
2031	18.1	20.9	–2.7	4.1	35,318	6,402	7,371	2.59	5.19	(584.25)	112.7	39,810
2032	18.2	21.1	–2.9	4.1	36,680	6,662	7,722	2.63	5.23	(604.69)	115.9	42,519
2033	18.2	21.2	–3.0	4.2	38,081	6,938	8,062	2.67	5.27	(607.14)	119.3	45,432
2034	18.3	21.3	–3.0	4.2	39,519	7,217	8,413	2.71	5.31	(610.82)	122.9	48,556
2035	18.3	21.4	–3.1	4.3	40,996	7,506	8,779	2.74	5.34	(614.50)	126.6	51,904
2036	18.4	21.6	–3.2	4.3	42,514	7,801	9,166	2.77	5.37	(622.61)	130.6	55,504
2037	18.4	21.7	–3.3	4.3	44,074	8,110	9,567	2.80	5.40	(628.07)	134.7	59,374
2038	18.4	21.8	–3.4	4.4	45,680	8,423	9,975	2.83	5.43	(631.58)	139.1	63,531
2039	18.5	22.0	–3.5	4.4	47,335	8,749	10,391	2.85	5.45	(631.12)	143.6	67,989

2040	18.5	22.1	−3.6	4.5	49,035	9,082	10,827	2.88	5.48	(633.61)	148.4	72,784
2041	18.6	22.2	−3.6	4.6	50,782	9,426	11,272	2.90	5.50	(632.63)	153.5	77,943
2042	18.6	22.3	−3.7	4.6	52,581	9,782	11,727	2.92	5.52	(629.12)	158.8	83,495
2043	18.7	22.4	−3.8	4.7	54,443	10,158	12,208	2.94	5.54	(625.67)	164.3	89,472
2044	18.7	22.5	−3.8	4.8	56,372	10,539	12,685	2.96	5.56	(618.12)	170.1	95,891
2045	18.7	22.6	−3.9	4.8	58,371	10,939	13,193	2.98	5.58	(612.45)	176.1	102,791
2046	18.8	22.7	−3.9	4.9	60,444	11,359	13,709	3.00	5.60	(602.20)	182.3	110,186
2047	18.8	22.7	−3.9	5.0	62,594	11,798	14,219	3.01	5.61	(585.45)	188.6	118,078
2048	18.9	22.8	−3.9	5.0	64,824	12,260	14,782	3.03	5.63	(575.36)	195.2	126,517
2049	19.0	22.8	−3.9	5.0	67,132	12,726	15,328	3.04	5.64	(559.89)	201.9	135,508
2050	19.0	22.9	−3.9	5.1	69,514	13,217	15,900	3.05	5.65	(544.63)	208.7	145,085
2051	19.1	22.9	−3.8	5.1	71,970	13,733	16,500	3.07	5.67	(529.80)	215.8	155,287
2052	19.1	23.0	−3.9	5.2	74,505	14,254	17,130	3.07	5.67	(521.22)	223.0	166,174
Total PV										(18,077)		30,109

Sources: Based on CBO projections and authors' calculations.

Note: Column 8 reports the discount rates used for spending and tax cash flows in that year, a 100 basis point increase relative to baseline. Column 4 reports the projected CBO's net interest cost over debt plus 100 basis points. Column 10 reports an upper bound on the PDV of projected primary surpluses in 2021 $ billions. Column 12 reports the debt dynamics: $D_{t+1} = D_t \times R_{t+1} + (T_{t+1} - G_{t+1})$. The variable R_{t+1} is taken from column 4.

bond yields rose by 176 basis points, the ten-year bond yield by 133 basis points, and the thirty-year bond yield by 131 basis points. We now explore what this shift in the term structure implies for our measure of fiscal capacity.

In our first exercise, we assume that this interest rate change only affects the rate at which we discount future surpluses but leaves future debt projections unchanged (as well as tax revenue, spending, and GDP projections).[34] The fiscal capacity bound becomes:

$$PV_{2021}^{upper}\left(\left\{S\right\}_{2022}^{2052}\right) + PV_{2021}^{upper}\left(D_{2052}\right) = -\$17.19 + \$22.78 \tag{18}$$

$$= \$5.59 \text{ trillion.}$$

We observe a substantial decline in fiscal capacity from the rise in interest rates, from \$12.38 trillion to \$5.59 trillion. At the new, higher rates, the valuation ratio of the GDP claim declines from 85.8 to 40.3. Servicing the same 185 percent debt/GDP after 2052 now requires annual surpluses of 4.59 percent of GDP, compared to 2.16 percent of GDP. Even though the fiscal adjustment after 2052 is more than twice as large, the fiscal capacity estimate falls by more than half.

Arguably, it is implausible that the CBO would not revise its interest rate forecast when projecting future debt service and future debt in light of these interest rate increases. To consider this additional effect, we add 156 basis points to the CBO's interest rate forecast in each year from 2022 to 2052. This 156 basis point increase is the increase in the five-year bond yield between December 31, 2021, and May 31, 2022, where the five-year maturity is chosen since it corresponds to the average maturity of the outstanding government bond portfolio. Under this assumption, the interest rate on the debt portfolio is 3.35 percent in 2022 and rises to 5.72 percent by 2052. We adjust the debt dynamics to account for the extra interest cost.

34. As in the previous exercise, increasing interest rates while keeping nominal GDP growth rates constant amounts to an increase in the real growth-adjusted return $r - g$. Such an increase in real rates is consistent with the data. The ten-year inflation-indexed Treasury bond yield increased from −1.04 percent on December 31, 2021, to +0.21 percent on May 31, 2022, an increase of 125 basis points. To do full-fledged counterfactual exercises, one would ideally like to use a general equilibrium model where GDP, inflation, interest rates, and fiscal policy are endogenously determined. A recent paper along these lines is Elenev and others (2021). Such a model would need to take a stance on what the fundamental shocks are that give rise to the changes in equilibrium interest rates: short-term or long-term productivity shocks, demand shocks, fiscal policy shocks, monetary policy shocks, etc. This is outside the scope of the current paper.

The debt in 2052 becomes \$187.5 trillion (251.6 percent of GDP) compared to \$137.9 trillion (185.0 percent of GDP) in the baseline. The upper bound on fiscal capacity becomes \$13.79 trillion, but that reflects the assumption that the surplus after 2052 now needs to be 6.24 percent per year compared to 4.59 percent in the previous exercise. In short, the fiscal capacity measure remains similar to the baseline value of \$12.38 trillion, but now the annual surpluses that need to be produced after 2052 are nearly triple what they were in the baseline. The massive change in required future fiscal adjustment reflects the high duration of the surplus claim at the end of 2021, when rates were very low, and the realization of a substantial increase in rates since then.

II.D. Debt Management

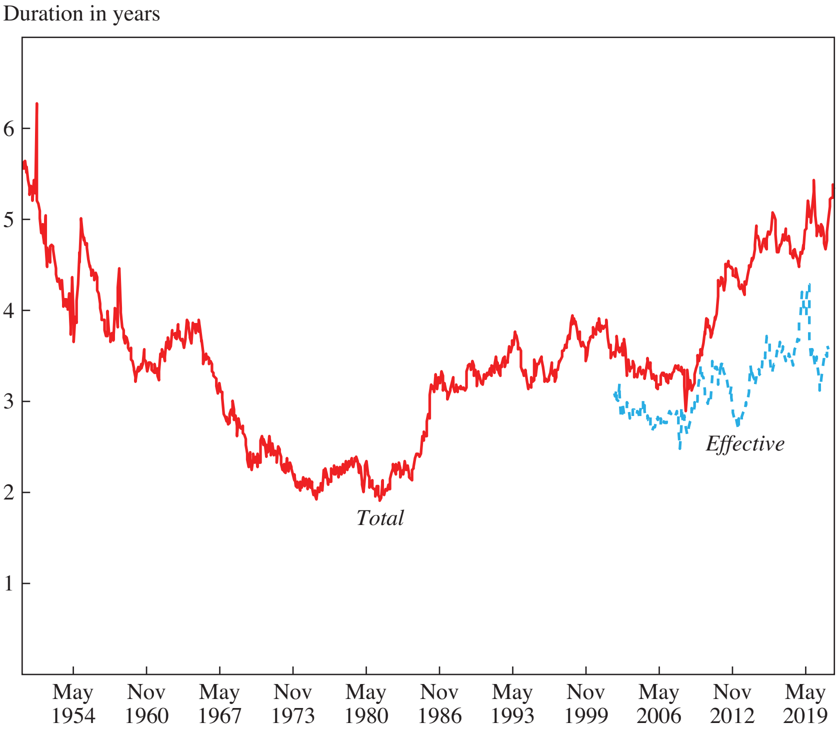
To eliminate duration risk, the Treasury would have to match the duration of its inflows to the duration of its outflows. The duration of the outstanding Treasuries is currently around five years, as shown in figure 4. In the baseline scenario, the US Treasury faces an extreme type of duration mismatch between its cash inflows (the surpluses) and cash outflows (the principal and coupon payments), a direct result of the back-loading of surpluses. This creates rollover risk and costly variation in future taxes and suggests that the Treasury should shift toward longer-maturity debt (Bhandari and others 2017).

In order to be fully hedged against interest rate risk, the Treasury should match the projected surplus (cash inflows) in each period to the coupon and principal payments (cash outflows), much like what a pension fund would typically try to do. To a first order, this requires matching the duration of the Treasury portfolio to the duration of the projected surpluses. In an optimal taxation framework, Bhandari and others (2017) show that the Ramsey planner wants to approximately match the duration of the projected surpluses, conditional on current tax rates, to the duration of the Treasury portfolio.

III. Adding Seigniorage from Convenience Yields

The United States is different from other countries because of its unique role as the world's safe asset supplier. Our calculations capture this by quantifying the seigniorage revenue from convenience yields. Our benchmark analysis abstracted from any convenience yields the Treasury earns on its sales of Treasuries. This section augments our baseline estimate of projected fiscal capacity with the present value of the revenue stream the government earns from convenience yields.

Figure 4. Duration of Treasuries Held by the Public



Source: Based on data from CRSP US Treasury Database. Copyright 2022 Center for Research in Security Prices (CRSP), the University of Chicago Booth School of Business.

As a result of being the world's safe asset supplier, the United States earns seigniorage revenue from its monopoly on the creation of safe, dollar-denominated assets. Jiang and others (2019) estimate that the United States earns around 60 basis points per annum in convenience yields on the entire US Treasury portfolio. The United States had a current debt/output ratio of 99.6 percent at the end of 2021. When the average convenience yield is 0.60 percent per annum, the Treasury collects 0.60 percent × 99.6 percent = 0.598 percent of GDP in convenience yield revenues per year.

Assumption 4: The seigniorage revenue on Treasuries is a constant fraction of GDP. This assumption of a constant seigniorage/GDP ratio implies that convenience yields decline as the debt/output ratio increases (to 185 percent of GDP in 2052 in the baseline model). Krishnamurthy and Vissing-Jorgensen (2012) provide evidence on downward-sloping demand curves for safe assets.[35] More recently, Mian, Straub, and Sufi (2021) analyze debt/output ratio dynamics in low interest rate environments when the government earns seigniorage from the convenience yields on government bonds, but faces a downward-sloping demand curve for liquidity and safety.

Table 4 reports the detailed calculations that account for convenience yields. Column 10 reports the seigniorage revenue in billions of dollars equal to 0.598 percent of GDP. Column 11 then discounts the seigniorage revenue back to 2021 dollars using the baseline discount rates. The sum of all this discounted seigniorage revenue between 2022 and 2052 is \$4.04 trillion in 2021 dollars. The upper bound on fiscal capacity is revised upward by this amount to \$16.4 trillion:

$$PV_{2021}^{upper}\left(\left\{T - G\right\}_{2022}^{2052}\right) + PV_{2021}^{upper}\left(D_{2052}\right) + PV_{2021}^{upper}\left(\left\{CS\right\}_{2022}^{2052}\right) = \$12.38 + \$4.04 = \$16.42 \text{ trillion.} \tag{19}$$

This number is still almost \$6 trillion short of the actual December 2021 value of government debt of \$22.28 trillion.

Under the assumption that seigniorage revenue continues to be a constant share of GDP after 2052, the government needs to run a smaller annual surplus of 1.56 percent (= 2.16 percent − 0.60 percent) of GDP after 2052, rather than 2.16 percent, to service the debt outstanding at the end of 2052. The smaller surpluses after 2052 also mean that the duration of the surplus claim is shorter than in the benchmark analysis.

Global investors may allocate additional borrowing capacity to the world's safe asset supplier, as argued by He, Krishnamurthy, and Milbradt (2019), not captured by the convenience yields. This may have been the case for the United Kingdom in the nineteenth century, but that privilege proved to be transitory (Chen and others 2022). While we cannot definitively rule out that the US government is one of the only countries to have permanently escaped the intertemporal budget constraint by engineering a bubble in the

35. In preference terms, if investors had utility defined over consumption and safe asset services, a constant expenditure share corresponds to an elasticity of substitution of one for the services provided by safe assets. The expenditure share accounted for by convenience yields is constant. Under the higher interest rate scenarios considered in the previous section, seigniorage revenue from convenience yields would be constant as a fraction of GDP even though convenience yields (seigniorage revenue divided by debt outstanding) would be falling as the debt/GDP ratio increased.

Table 4. Fiscal Capacity with Convenience Yields

Year	T/Y (%) (1)	G/Y (%) (2)	Y ($ billions) (3)	T ($ billions) (4)	G ($ billions) (5)	$y^{\$}$ (%) (6)	$r^{\$,y}$ (%) (7)	$PV(T-G)$ (8)	D/Y (9)	CS ($ billions) (10)	$PV(CS)$ (11)
2022	19.6	21.9	24,694	4,836	5,405	0.42	3.02	(552.26)	97.9	147.63	143.30
2023	18.6	20.7	26,240	4,889	5,419	0.76	3.36	(495.70)	96.0	156.87	146.85
2024	18.0	20.3	27,291	4,924	5,535	0.99	3.59	(549.75)	96.1	163.15	146.79
2025	17.6	20.1	28,271	4,982	5,696	1.15	3.75	(616.35)	97.5	169.01	145.87
2026	18.0	20.4	29,266	5,280	5,962	1.27	3.87	(564.72)	98.8	174.97	144.71
2027	18.3	20.4	30,332	5,548	6,201	1.36	3.96	(517.27)	100.0	181.33	143.63
2028	18.2	20.6	31,487	5,716	6,486	1.43	4.03	(583.70)	102.0	188.24	142.72
2029	18.1	20.7	32,716	5,934	6,773	1.49	4.09	(608.55)	103.2	195.59	141.89
2030	18.1	20.8	33,996	6,161	7,066	1.55	4.15	(627.67)	105.3	203.24	141.02
2031	18.1	20.9	35,318	6,402	7,371	1.59	4.19	(642.81)	107.5	211.14	140.05
2032	18.2	21.1	36,680	6,662	7,722	1.63	4.23	(671.66)	109.6	219.28	138.99
2033	18.2	21.2	38,081	6,938	8,062	1.67	4.27	(680.82)	112.0	227.66	137.83
2034	18.3	21.3	39,519	7,217	8,413	1.71	4.31	(691.49)	114.4	236.26	136.57
2035	18.3	21.4	40,996	7,506	8,779	1.74	4.34	(702.29)	117.0	245.09	135.22
2036	18.4	21.6	42,514	7,801	9,166	1.77	4.37	(718.35)	119.8	254.16	133.79
2037	18.4	21.7	44,074	8,110	9,567	1.80	4.40	(731.56)	122.7	263.49	132.29
2038	18.4	21.8	45,680	8,423	9,975	1.83	4.43	(742.66)	125.8	273.09	130.74
2039	18.5	22.0	47,335	8,749	10,391	1.85	4.45	(749.19)	129.1	282.98	129.15

2040	18.5	22.1	49,035	9,082	10,827	1.88	4.48	(759.32)	132.5	293.15	127.51
2041	18.6	22.2	50,782	9,426	11,272	1.90	4.50	(765.37)	136.1	303.59	125.84
2042	18.6	22.3	52,581	9,782	11,727	1.92	4.52	(768.38)	139.9	314.35	124.15
2043	18.7	22.4	54,443	10,158	12,208	1.94	4.54	(771.44)	143.9	325.48	122.46
2044	18.7	22.5	56,372	10,539	12,685	1.96	4.56	(769.40)	148.0	337.01	120.79
2045	18.7	22.6	58,371	10,939	13,193	1.98	4.58	(769.60)	152.3	348.96	119.13
2046	18.8	22.7	60,444	11,359	13,709	2.00	4.60	(763.93)	156.7	361.36	117.49
2047	18.8	22.7	62,594	11,798	14,219	2.01	4.61	(749.74)	161.2	374.21	115.88
2048	18.9	22.8	64,824	12,260	14,782	2.03	4.63	(743.84)	165.8	387.54	114.29
2049	19.0	22.8	67,132	12,726	15,328	2.04	4.64	(730.74)	170.5	401.34	112.71
2050	19.0	22.9	69,514	13,217	15,900	2.05	4.65	(717.59)	175.2	415.58	111.14
2051	19.1	22.9	71,970	13,733	16,500	2.07	4.67	(704.70)	180.1	430.26	109.57
2052	19.1	23.0	74,505	14,254	17,130	2.07	4.67	(699.91)	185.0	445.42	108.37
Total								(21,161)	33,540		4,041

Sources: Based on CBO projections released May 2022 and authors' calculations.

Note: Column 10 reports an estimate of the seigniorage revenue collected by the Treasury. We use a convenience yield of 60 basis points per annum. PDV of projected surpluses (column 8) and seigniorage (column 11) are measured at the end of 2021.

bond market, it seems prudent to assume that this is not the case, especially from the perspective of future US generations.

III.A. Broad and Narrow Convenience Yields

In our analysis above, we kept the rate used to discount future surpluses and future debt unchanged when introducing convenience yields. Implicitly, this assumed that there was a decline in the risk premium (of 60 basis points) that exactly offset the implied increase in the true risk-free yield (of 60 basis points). Jiang and others (2019) refer to this as a narrow convenience yield—a convenience yield that does not accrue to asset classes other than Treasuries. By not increasing the discount rate when the true risk-free rate increased, we did not decrease the present value of the seigniorage revenue from convenience yields as well as the present value of primary surpluses. If anything, this overstated the extra fiscal capacity that convenience yields generated. Since we showed that this generous upper bound on fiscal capacity inclusive of convenience yields is still too low, our results are conservative.

Recently, Reis (2021) has convincingly argued that convenience yields on US Treasuries could be much larger than 60 basis points per year. While larger convenience yields generate an additional source of revenue that expands fiscal capacity, they also generate a discount rate effect that shrinks fiscal capacity. The reason is that large convenience yields are likely broad convenience yields, which apply to assets beyond US Treasuries. Such broad convenience yields raise the true risk-free interest rate (on risk-free assets without convenience) but also the discount rate on risky assets such as the GDP claim. Risk premia declines do not fully offset the risk-free rate effect. Higher discount rates lower the present value of the seigniorage revenue stream and the primary surplus stream, all else equal. Hence, it is not clear that even much larger convenience yields actually result in more fiscal capacity.

IV. Front-Loaded Fiscal Adjustment

So far, we have established that the current level of debt is higher than our upper bound on fiscal capacity, even after including seigniorage revenue from convenience yields. This raises the question how the US economy can increase its fiscal capacity. A natural answer is that it must increase its surpluses.

This section implements a counterfactual exercise by asking by how much CBO primary surplus projections have to rise in order to obtain a fiscal capacity estimate consistent with the 99.7 percent debt/output ratio at the end of 2021. We consider level shifts that raise the surplus/GDP ratio in each year from 2022 until 2052. This policy change also affects the debt dynamics. We compute these projected debt dynamics, $D_{t+1} = D_t \times R_{t+1} + S_{t+1}$, using the new projected primary surpluses and the CBO's interest rate projections. When performing this counterfactual, we make the following assumption.

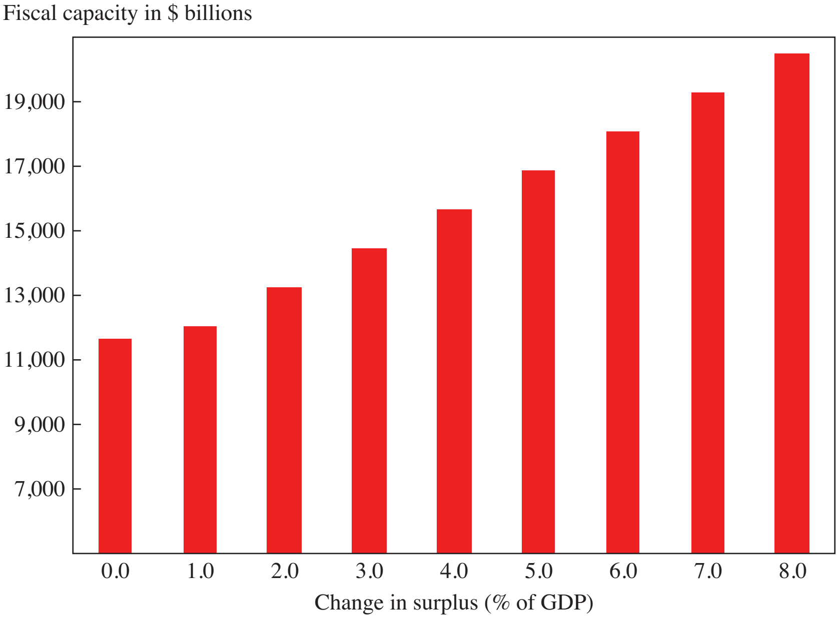
Assumption 5: We assume the surplus changes relative to the CBO baseline do not change the projected growth rate of GDP nor the yield curve. We first consider an increase in the primary surplus by 3 percentage points of GDP in each of the years between 2022 and 2052 relative to the CBO projection. This fiscal adjustment increases the PDV of surpluses between 2022 and 2052 from −\$21.16 trillion in the baseline to −\$0.88 trillion. Hence, a fiscal adjustment of 3 percent per year nearly eliminates all deficits over the next thirty-one years in present value. The higher primary surpluses decrease the value of debt outstanding at the end of 2052 to 87.5 percent of GDP. Discounted back to 2021, that is \$15.86 trillion. Combined, this raises the upper bound on fiscal capacity from \$12.38 trillion in the benchmark to \$14.97 trillion:

$$\begin{aligned} PV_{2021}^{upper}\left(\left\{T-G\right\}_{2022}^{2052}\right) + PV_{2021}^{upper}\left(D_{2052}\right) &= -\$0.88 + \$15.86 \\ &= \$14.97 \text{ trillion.} \end{aligned} \tag{20}$$

In this counterfactual exercise, the US Treasury front-loads the fiscal adjustment, compared to the benchmark case in which the government waits until after 2052 before running primary surpluses. In this front-loaded case, the United States only needs a 1.02 percent annual primary surplus after 2052, less than half the 2.16 percent annual surplus number in the baseline. Figure 2 plots this front-loaded path of surpluses. In this scenario, the duration of the surplus claim declines to 126 years from 283 years in the baseline.

Next, we repeat the projected fiscal capacity calculation assuming increases in the surplus/GDP ratio in each of the years between 2022 and 2052 relative to the CBO projection ranging from 0 percent per year (baseline) to 8 percent per year in 1 percentage point increments. Figure 5 plots the projected fiscal capacity on the y-axis against the increase in the projected surplus/GDP ratio for the period 2022–2052. The previous example of a 3 percent increase lies in the middle of this graph.

Figure 5. Fiscal Capacity for Additional Surpluses in 2022–2052



Source: Authors' calculations.
Note: Change in primary surplus as a percentage of US GDP in each year between 2022 and 2052 is relative to the baseline CBO projection.

To get to an upper bound on fiscal capacity of \$18.3 trillion, we need an extra primary surplus of 6 percent of GDP in all years between 2022 and 2052. Table 5 provides all of the details of the calculation. This scenario pushes the debt/GDP ratio into negative territory by 2050. The fiscal capacity bound reaches:

$$PV_{2021}^{upper}\left(\left\{T - G\right\}_{2022}^{2052}\right) + PV_{2021}^{upper}\left(D_{2052}\right) = \$19.39 - \$1.09 \tag{21}$$

$$= \$18.30 \text{ trillion.}$$

Once we factor in the \$4.04 trillion in convenience yield revenues, this scenario of 6 percent additional surpluses between 2022–2052 produces a fiscal capacity estimate that essentially matches the observed debt outstanding of \$22.28 trillion as of the end of 2021. Figure 1 also plots this 6 percent extra surpluses path of completely front-loaded surpluses. In this scenario, the government can run a small primary deficit of 0.07 percent of GDP in each year after 2052.

Figure 6 plots the contribution of each surplus cash flow to the overall duration of the surplus claim in this front-loaded scenario with 6 percent additional surpluses. This surplus claim has a duration of 6.95 years, which is close to that of the outstanding Treasury portfolio.[36] In sum, if the government wants to match the duration of the surpluses (cash inflows) to the duration of the outstanding portfolio of Treasury debt (cash outflows), it needs to raise annual surpluses relative to the CBO scenario by about 6 percent per year over the next thirty-one years. Suffice to say that this is a massive fiscal effort.

V. Countercyclical Tax Regime

Can the United States run steady-state deficits and maintain fiscal capacity, as many have claimed? Not according to standard finance, unless the US federal government changes the fiscal regime from countercyclical to pro-cyclical. The US Treasury would have to render the tax claim less risky than the spending claim. Only in that case would our upper bound calculation fail, because assumption 1 above fails. In this case, the US taxpayers would be providing insurance to bondholders (Jiang and others 2020). This insurance premium would allow the United States to run steady-state deficits.

Hence, the only way to reconcile the CBO projections with the value of US Treasuries is to use a much lower discount rate for the tax cash flows than for the spending cash flows. Importantly, this is necessary if we want the entire debt to be zero beta or risk-free. However, this condition is not satisfied in postwar US data because of the pro-cyclical nature of tax revenue and the countercyclical nature of spending (Jiang and others 2019). We explore this hypothetical scenario, but we emphasize that we do not think this regime shift is either likely or desirable.

If the US government were to radically change its future fiscal policy and raise more tax revenue as a share of GDP in recessions, this would make the tax claim less risky than the spending claim. We entertain this possibility because this regime change can sustain (modest) steady-state deficits. In this regime, taxpayers and transfer recipients provide insurance against business cycle risk to the bondholders. Taxpayers pay more taxes

36. The duration is sensitive to the additional surplus. Raising the additional surplus from 6.0 percent to 6.1 percent per year until 2052 lowers the duration from 6.95 to 3.45 years.

Table 5. Fiscal Capacity with 6 Percent Extra Surplus in 2022–2052

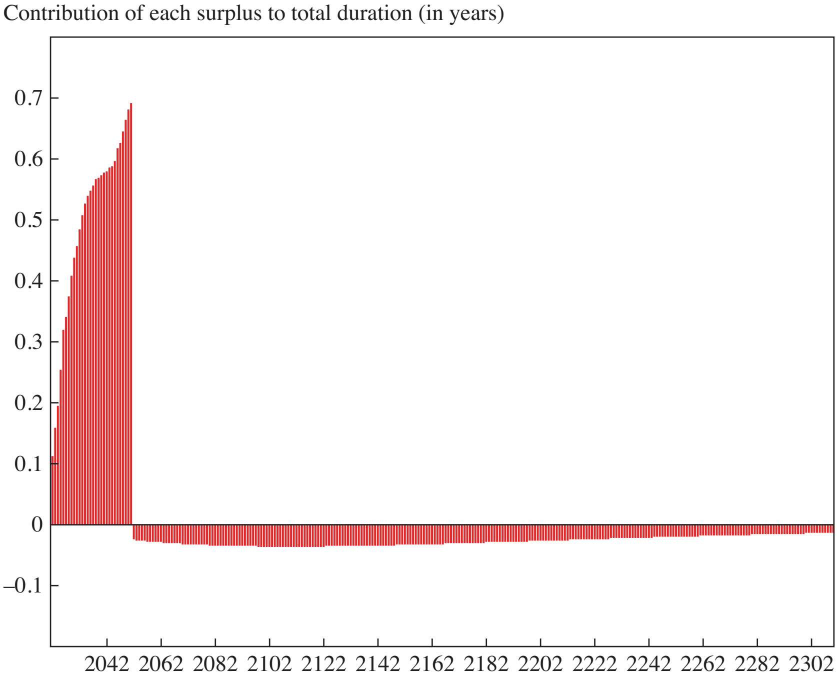
Year	T/Y (%) (1)	G/Y (%) (2)	$(T-G)/Y$ (%) (3)	NI/D (%) (4)	Y ($ billions) (5)	T ($ billions) (6)	G ($ billions) (7)	$y^{\$}$ (%) (8)	$r^{\$,y}$ (%) (9)	$PV(T-G)$ ($ billions) (10)	D/Y (11)	D ($ billions) (12)
2022	25.6	21.9	3.7	1.8	24,694	6,318	5,405	0.42	3.02	885.92	88.2	21,770
2023	24.6	20.7	4.0	1.8	26,240	6,464	5,419	0.76	3.36	978.13	80.5	21,124
2024	24.0	20.3	3.8	2.1	27,291	6,561	5,535	0.99	3.59	923.46	75.3	20,538
2025	23.6	20.1	3.5	2.3	28,271	6,678	5,696	1.15	3.75	847.66	70.8	20,029
2026	24.0	20.4	3.7	2.5	29,266	7,036	5,962	1.27	3.87	887.63	66.5	19,450
2027	24.3	20.4	3.8	2.6	30,332	7,368	6,201	1.36	3.96	924.26	62.0	18,792
2028	24.2	20.6	3.6	2.8	31,487	7,605	6,486	1.43	4.03	848.70	57.8	18,195
2029	24.1	20.7	3.4	2.9	32,716	7,897	6,773	1.49	4.09	815.51	53.8	17,595
2030	24.1	20.8	3.3	3.0	33,996	8,201	7,066	1.55	4.15	787.59	50.0	16,985
2031	24.1	20.9	3.3	3.1	35,318	8,521	7,371	1.59	4.19	762.76	46.3	16,356
2032	24.2	21.1	3.1	3.1	36,680	8,863	7,722	1.63	4.23	723.27	42.9	15,729
2033	24.2	21.2	3.0	3.2	38,081	9,223	8,062	1.67	4.27	702.49	39.6	15,073
2034	24.3	21.3	3.0	3.2	39,519	9,588	8,413	1.71	4.31	679.15	36.4	14,387
2035	24.3	21.4	2.9	3.3	40,996	9,965	8,779	1.74	4.34	654.79	33.3	13,671
2036	24.4	21.6	2.8	3.3	42,514	10,352	9,166	1.77	4.37	624.36	30.4	12,937
2037	24.4	21.7	2.7	3.3	44,074	10,754	9,567	1.80	4.40	596.13	27.6	12,182
2038	24.4	21.8	2.6	3.4	45,680	11,164	9,975	1.83	4.43	569.46	25.0	11,406
2039	24.5	22.0	2.5	3.4	47,335	11,589	10,391	1.85	4.45	546.99	22.4	10,599

2040	24.5	22.1	2.4	3.5	49,035	12,024	10,827	1.88	4.48	520.44	19.9	9,772
2041	24.6	22.2	2.4	3.6	50,782	12,473	11,272	1.90	4.50	497.62	17.6	8,918
2042	24.6	22.3	2.3	3.6	52,581	12,937	11,727	1.92	4.52	477.64	15.3	8,033
2043	24.7	22.4	2.2	3.7	54,443	13,424	12,208	1.94	4.54	457.62	13.1	7,114
2044	24.7	22.5	2.2	3.8	56,372	13,921	12,685	1.96	4.56	442.89	10.9	6,147
2045	24.7	22.6	2.1	3.8	58,371	14,441	13,193	1.98	4.58	426.05	8.8	5,135
2046	24.8	22.7	2.1	3.9	60,444	14,986	13,709	2.00	4.60	415.28	6.7	4,058
2047	24.8	22.7	2.1	4.0	62,594	15,553	14,219	2.01	4.61	413.25	4.6	2,885
2048	24.9	22.8	2.1	4.0	64,824	16,150	14,782	2.03	4.63	403.18	2.5	1,633
2049	25.0	22.8	2.1	4.0	67,132	16,754	15,328	2.04	4.64	400.44	0.4	274
2050	25.0	22.9	2.1	4.1	69,514	17,388	15,900	2.05	4.65	397.83	–1.7	(1,203)
2051	25.1	22.9	2.2	4.1	71,970	18,051	16,500	2.07	4.67	394.96	–3.9	(2,803)
2052	25.1	23.0	2.1	4.2	74,505	18,724	17,130	2.07	4.67	387.75	–6.1	(4,513)
Total PV										19,393		(1,098)

Source: Based on CBO primary surplus projections plus an additional 6 percent of GDP in primary surplus for each year from 2022 until 2052.

Note: Column 8 reports the discount rates used for spending and tax cash flows in that year. Column 4 reports projected net interest cost over debt. Column 10 reports an upper bound on the PDV of projected primary surpluses in 2021 \$ billions. Column 12 reports the debt dynamics: $D_{t+1} = D_t \times R_{t+1} + (T_{t+1} - G_{t+1})$, where R_{t+1} is taken from column 4.

Figure 6. Duration Composition in Front-Loaded Scenario



Source: Authors' calculations.

Note: Contribution of each payment $\frac{k \times PV(S_{2021+k})}{\sum_{h=1} PV(S_{2021+h})}$ to the total duration. Primary surplus: CBO baseline projection plus 6 percent of GDP in each year between 2022 and 2052. The duration (in years) is the sum of the bars shown.

as a fraction of GDP in recessions, while transfer recipients receive less. To make this concrete, when taxpayers wake up in a recession, the CBO should be projecting larger tax revenue as a fraction of GDP in PDV, and smaller spending as a fraction of GDP than in an expansion, meaning that the bottom row of column 9 in table 1 increases (decreases) when a recession (expansion) starts.

In the steady-state, the valuation of future surpluses is given the price/dividend ratio on a claim to GDP times the steady-state surplus:

$$(22) \qquad PV_{2021}^{upper}\left(\{T - G\}_{2052}^{\infty}\right) = \sum_{h=1}^{\infty} \frac{T_{2021+h}}{\left(1 + r^{\$,t}(h)\right)^h} - \sum_{h=1}^{\infty} \frac{G_{2021+h}}{\left(1 + r^{\$,g}(h)\right)^h}$$

$$= \left(pd^t \times \frac{T}{Y} - pd^g \times \frac{G}{Y}\right) \times Y_{2021}.$$

Table 6. US Treasury Balance Sheet in Steady-State Countercyclical Tax Example

Assets		*Liabilities*	
$PV_{2021}(\{T\})/Y_{2021}$	$19.71 = 18.71\% \times 105.3$	$PV_{2021}(\{G\})/Y_{2021}$	$18.79 = 21.9\% \times 85.8$
		D/Y_{2021}	$0.99 = 18.71\% \times 105.3 - 21.9\% \times 85.8$
Total	19.71	Total	19.71

Source: Authors' calculations.

Note: Market values are expressed as a multiple of US GDP at the end of 2021. The steady-state example is based on the actual spending/GDP ratio in 2022. In this example, the risk premium on the tax claim is 22 basis points lower than the risk premium on the spending claim.

If the tax claim is less risky, and the price/dividend ratio on the tax claim exceeds that on the spending claim, $pd^g < pd^t$, then a steady-state deficit is consistent with positive fiscal capacity. Table 6 provides a simple example, starting from the actual spending/output ratio for 2022. If the multiple on the tax claim is boosted to 105.3, then the US government can run a steady-state deficit of 3.19 percent, the CBO-projected average deficit. The implied debt/output ratio is still 0.99. The government can engineer this outcome by committing to a pro-cyclical fiscal policy (leaning with the wind) that raises taxes T/Y in bad times, thus lowering the risk premium. However, this is not a free lunch. Taxpayers are being asked to bear more business cycle risk in order to provide insurance to bondholders, allowing the government to earn an insurance premium each year that is 3.19 percent of GDP. This is counterfactual. In advanced economies, it is the government that typically provides insurance against business cycle risk.[37]

Let's turn to the detailed CBO projections. Suppose that the tax claim's appropriate discount rate is 100 basis points lower than the discount rate for the output claim. Table 7 reports the calculations. Now the sum of (the upper bound on) the PDV of the tax revenue minus spending cash flows from 2022 to 2052 adds up to −\$846 billion:

$$(23)\quad PV_{2021}^{upper}\left(\left\{T - G\right\}_{2052}^{2052}\right) = \sum_{h=1}^{31} \frac{T_{2021+j}}{\left(1 + r^{\$,y}\left(h\right) - 0.01\right)^{h}} - \sum_{h=1}^{31} \frac{G_{2021+j}}{\left(1 + r^{\$,y}\left(h\right)\right)^{h}}$$

$$= \$0.85 \text{ trillion.}$$

The lower discount rate for the tax revenue claim expands our estimate of fiscal capacity. In this case, the total PDV of deficits, computed

37. See Jiang and others (2020) for evidence on the GDP growth betas of US taxes and spending over longer horizons. They find large positive GDP growth betas for taxes at shorter horizons, and negative GDP growth betas for spending.

Table 7. Fiscal Capacity with Countercyclical Tax Revenues

Year	T/Y (%) (1)	G/Y (%) (2)	$(T-G)/Y$ (%) (3)	Y ($ billions) (4)	T ($ billions) (5)	G ($ billions) (6)	$y^{\$}$ (%) (7)	$r^{\$,y}$ (%) (8)	$PV(T)$ ($ billions) (9)	$PV(G)$ ($ billions) (10)	D/Y (11)	D ($ billions) (12)
2022	19.6	21.9	−2.3	24,694	4,836	5,405	0.42	3.02	4,740.22	5,246.47	97.9	24,173
2023	18.6	20.7	−2.0	26,240	4,889	5,419	0.76	3.36	4,667.08	5,072.90	96.0	25,193
2024	18.0	20.3	−2.2	27,291	4,924	5,535	0.99	3.59	4,560.75	4,979.69	96.1	26,217
2025	17.6	20.1	−2.5	28,271	4,982	5,696	1.15	3.75	4,469.39	4,915.89	97.5	27,561
2026	18.0	20.4	−2.3	29,266	5,280	5,962	1.27	3.87	4,583.17	4,931.48	98.8	28,925
2027	18.3	20.4	−2.2	30,332	5,548	6,201	1.36	3.96	4,657.12	4,911.99	100.0	30,326
2028	18.2	20.6	−2.4	31,487	5,716	6,486	1.43	4.03	4,636.87	4,917.44	102.0	32,105
2029	18.1	20.7	−2.6	32,716	5,934	6,773	1.49	4.09	4,650.55	4,913.48	103.2	33,760
2030	18.1	20.8	−2.7	33,996	6,161	7,066	1.55	4.15	4,662.85	4,902.70	105.3	35,808
2031	18.1	20.9	−2.7	35,318	6,402	7,371	1.59	4.19	4,676.14	4,889.04	107.5	37,949
2032	18.2	21.1	−2.9	36,680	6,662	7,722	1.63	4.23	4,695.09	4,894.36	109.6	40,213
2033	18.2	21.2	−3.0	38,081	6,938	8,062	1.67	4.27	4,715.35	4,881.23	112.0	42,636
2034	18.3	21.3	−3.0	39,519	7,217	8,413	1.71	4.31	4,728.10	4,863.04	114.4	45,219
2035	18.3	21.4	−3.1	40,996	7,506	8,779	1.74	4.34	4,738.61	4,843.22	117.0	47,975
2036	18.4	21.6	−3.2	42,514	7,801	9,166	1.77	4.37	4,744.45	4,824.82	119.8	50,926
2037	18.4	21.7	−3.3	44,074	8,110	9,567	1.80	4.40	4,749.69	4,803.35	122.7	54,088
2038	18.4	21.8	−3.4	45,680	8,423	9,975	1.83	4.43	4,749.26	4,775.27	125.8	57,472
2039	18.5	22.0	−3.5	47,335	8,749	10,391	1.85	4.45	4,748.01	4,742.30	129.1	61,087

2040	18.5	22.1	−3.6	49,035	9,082	10,827	1.88	4.48	4,742.43	4,709.73	132.5	64,963
2041	18.6	22.2	−3.6	50,782	9,426	11,272	1.90	4.50	4,735.76	4,672.62	136.1	69,115
2042	18.6	22.3	−3.7	52,581	9,782	11,727	1.92	4.52	4,727.50	4,631.66	139.9	73,568
2043	18.7	22.4	−3.8	54,443	10,158	12,208	1.94	4.54	4,721.94	4,593.42	143.9	78,343
2044	18.7	22.5	−3.8	56,372	10,539	12,685	1.96	4.56	4,711.61	4,546.68	148.0	83,447
2045	18.7	22.6	−3.9	58,371	10,939	13,193	1.98	4.58	4,702.92	4,504.00	152.3	88,909
2046	18.8	22.7	−3.9	60,444	11,359	13,709	2.00	4.60	4,696.07	4,457.38	156.7	94,734
2047	18.8	22.7	−3.9	62,594	11,798	14,219	2.01	4.61	4,689.75	4,403.10	161.2	100,911
2048	18.9	22.8	−3.9	64,824	12,260	14,782	2.03	4.63	4,685.91	4,359.44	165.8	107,481
2049	19.0	22.8	−3.9	67,132	12,726	15,328	2.04	4.64	4,676.48	4,304.70	170.5	114,436
2050	19.0	22.9	−3.9	69,514	13,217	15,900	2.05	4.65	4,669.42	4,252.18	175.2	121,798
2051	19.1	22.9	−3.8	71,970	13,733	16,500	2.07	4.67	4,664.32	4,201.80	180.1	129,588
2052	19.1	23.0	−3.9	74,505	14,254	17,130	2.07	4.67	4,670.11	4,167.90	185.0	137,852
Total										(846.4)	33,540	

Sources: Based on CBO projections released May 2022 and authors' calculations.

Note: Columns 9 and 10 report the PDV of tax revenue and spending, respectively. The discount rate used for the tax claim is 100 basis points lower than that used for the spending claim.

as the difference between the sum of columns 9 and 10, has shrunk from \$21.16 trillion to \$ 0.85 trillion. If we combine this with the \$33.54 trillion in PDV of future debt, we end up with a total value of \$32.7 trillion for the value of debt at the end of 2021.

$$(24) \qquad PV_{2021}^{upper}\left(\left\{T - G\right\}_{2052}^{2052}\right) + PV_{2021}^{upper}\left(D_{2052}\right) = \$32.69 \text{ trillion.}$$

This measure of projected fiscal capacity comfortably exceeds the current debt outstanding at the end of 2021. This exercise goes to show that the nature of risk in tax revenues (and government spending) is crucial for the magnitude of the projected fiscal capacity. A radical fiscal regime shift of the kind entertained in this section, where tax rates go up in recessions, seems unlikely because of the pain it would inflict on taxpayers.

VI. Conclusion

We develop a new approach based on textbook finance to assess the government's projected fiscal capacity, and we apply this framework to the CBO's projections of the federal government's primary surpluses. Using plausible discount rate assumptions, we measure the fiscal capacity of the US federal government implied by the May 2022 CBO projections. In spite of the historically low interest rates at the end of 2021, the upper bound on fiscal capacity is only around 56 percent of the observed debt outstanding in 2021.

From the vantage point of standard, neoclassical finance, our findings would imply that the Treasury market has likely priced in a large fiscal correction relative to the CBO baseline projections. In this scenario, future surpluses will increase to close the gap. However, we cannot rule out that Treasuries are mispriced. Treasury investors may be optimistic about future surpluses or they may fail to price in future inflation. In this case, bond yields will need to increase to close the gap.

Many authors have emphasized that low rates create additional fiscal capacity for the United States, but they have ignored the impact of low rates on the risk of future fiscal adjustment due to the duration mismatch. The back-loading of surpluses creates a large duration mismatch between the government's assets, its future surpluses, and its liabilities, its promised coupon and principal payments on the Treasury portfolio. Because of the back-loading of future surpluses, the Treasury faces a duration mismatch between its cash inflows and outflows. Modest increases in interest rates, of the kind the US economy experienced in the first half of 2022, then lead to sharp increases in the size of required fiscal adjustments.

Our analysis highlights a shortcoming in the standard fiscal sustainability analysis, namely, the practice of discounting future primary surpluses and future debt at the risk-free interest rate to measure fiscal capacity. This standard practice ignores a basic insight from finance that the discount rate should always reflect the risk of the cash flows. Fiscal cash flow projections are always made relative to GDP projections. But the future course of the economy is unknown, and hence fundamentally risky. Future primary surpluses inherit the risk in future GDP and are at least as risky as future GDP unless the government chooses countercyclical primary surpluses. Hence, future surpluses should be discounted at a rate that includes a risk premium that is at least as large as the GDP risk premium.

To be clear, there is considerable uncertainty about the GDP risk premium. Our baseline estimate for the total wealth valuation multiple is 85. A lower risk premium and a higher multiple leads to higher estimates of fiscal capacity, but this would imply counterfactual valuation multiples well in excess of 100 for unlevered companies growing at the same rate as the US economy. Lower discount rates also lead to an even larger duration mismatch between the government's assets and liabilities, and hence even larger fiscal vulnerability to the risk of rising interest rates. Model uncertainty is not a panacea to get us out of the fiscal conundrum.

ACKNOWLEDGMENTS We gratefully acknowledge financial support from the National Science Foundation. We are grateful to Andrew Atkeson for suggesting this exercise to us. Peter DeMarzo, Janice Eberly, William Gale, Deborah Lucas, and Phillip Swagel provided very helpful and detailed comments on a first draft. We are grateful to the *BPEA* conference participants for helpful comments.

References

Acalin, Julien, and Laurence Ball. 2022. "Did the U.S. Really Grow Its Way Out of Its WWII Debt?" Working Paper. https://julienacalin.com/publication/workinprogress/us-debt-after-wwii.

Aguiar, Mark A., Manuel Amador, and Cristina Arellano. 2021. "Micro Risks and Pareto Improving Policies with Low Interest Rates." Working Paper 28996. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28996.

Andersen, Torben G., Nicola Fusari, and Viktor Todorov. 2015. "The Risk Premia Embedded in Index Options." *Journal of Financial Economics* 117, no. 3: 558–84.

Andolfatto, David. 2020. "Does the National Debt Matter?" *Regional Economist*, Federal Reserve Bank of St. Louis. https://www.stlouisfed.org/publications/regional-economist/fourth-quarter-2020/does-national-debt-matter.

Backus, David, Nina Boyarchenko, and Mikhail Chernov. 2018. "Term Structures of Asset Prices and Returns." *Journal of Financial Economics* 129, no. 1: 1–23.

Ball, Laurence M., and N. Gregory Mankiw. 2021. "Market Power in Neoclassical Growth Models." Working Paper 28538. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28538.

Bansal, Ravi, and Amir Yaron. 2004. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance* 59, no. 4: 1481–509.

Barro, Robert J. 2020. "*r* Minus *g*." Working Paper 28002. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28002.

Bhandari, Anmol, David Evans, Mikhail Golosov, and Thomas J. Sargent. 2017. "The Optimal Maturity of Government Debt." Working Paper. https://bfi.uchicago.edu/wp-content/uploads/1_Evans-Paper.pdf.

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

Bohn, Henning. 1998. "The Behavior of U.S. Public Debt and Deficits." *Quarterly Journal of Economics* 113, no. 3: 949–63.

Botev, Jarmila, Jean-Marc Fournier, and Annabelle Mourougane. 2016. "A Re-assessment of Fiscal Space in OECD Countries." Working Paper 1352. Paris: Organisation for Economic Cooperation and Development. https://www.oecd-ilibrary.org/economics/a-re-assessment-of-fiscal-space-in-oecd-countries_fec60e1b-en.

Brunnermeier, Markus K., Sebastian A. Merkel, and Yuliy Sannikov. 2022. "Debt as Safe Asset." Working Paper 29626. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29626.

Chen, Zefeng, Zhengyang Jiang, Hanno Lustig, Stijn Van Nieuwerburgh, and Mindy Z. Xiaolan. 2022. "Exorbitant Privilege Gained and Lost: Fiscal Implications." Working Paper 30059. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30059.

Congressional Budget Office (CBO). 2021a. *The 2021 Long-Term Budget Outlook*. https://www.cbo.gov/publication/57038.

Congressional Budget Office (CBO). 2021b. *The Budget and Economic Outlook: 2021 to 2031*. https://www.cbo.gov/publication/56970.

Congressional Budget Office (CBO). 2021c. *An Update to the Budget and Economic Outlook: 2021 to 2031*. https://www.cbo.gov/publication/57218.

Congressional Budget Office (CBO). 2022. *The Budget and Economic Outlook: 2022 to 2032*. https://www.cbo.gov/publication/57950.

D'Amico, Stefania, William English, David López-Salido, and Edward Nelson. 2012. "The Federal Reserve's Large-Scale Asset Purchase Programmes: Rationale and Effects." *Economic Journal* 122, no. 564: F415–46.

Dumas, Bernard, Paul Ehling, and Chunyu Yang. 2021. "The Debt Capacity of a Government." Working Paper 29434. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29434.

Elenev, Vadim, Tim Landvoigt, Patrick J. Shultz, and Stijn Van Nieuwerburgh. 2021. "Can Monetary Policy Create Fiscal Capacity?" Working Paper 29129. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29129.

Farhi, Emmanuel, and François Gourio. 2018. "Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia." *Brookings Papers on Economic Activity*, Fall, 147–223.

Fernandez, Pablo, Sofia Bañuls, and Pablo Fernancez Acin. 2021. "Survey: Market Risk Premium and Risk-Free Rate Used for 88 Countries in 2021." Working Paper. Social Science Research Network, June 16. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3861152.

Furman, Jason, and Lawrence Summers. 2020. "A Reconsideration of Fiscal Policy in the Era of Low Interest Rates." Discussion Paper. https://www.brookings.edu/wp-content/uploads/2020/11/furman-summers-fiscal-reconsideration-discussion-draft.pdf.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright. 2007. "The U.S. Treasury Yield Curve: 1961 to the Present." *Journal of Monetary Economics* 54, no. 8: 2291–304.

Hall, George J., and Thomas J. Sargent. 2011. "Interest Rate Risk and Other Determinants of Post-WWII US Government Debt/GDP Dynamics." *American Economic Journal: Macroeconomics* 3, no. 3: 192–214.

Hansen, Lars Peter, William Roberds, and Thomas J. Sargent. 1991. "Time Series Implications of Present Value Budget Balance and of Martingale Models of Consumption and Taxes." In *Rational Expectations Econometrics*, edited by Lars Peter Hansen and Thomas J. Sargent. Boulder, Colo.: Westview Press.

He, Zhiguo, Arvind Krishnamurthy, and Konstantin Milbradt. 2019. "A Model of Safe Asset Determination." *American Economic Review* 109, no. 4: 1230–62.

Jensen, Michael C. 1972. "Capital Markets: Theory and Evidence." *Bell Journal of Economics and Management Science* 3, no. 2: 357–98.

Jiang, Zhengyang, Hanno Lustig, Stijn Van Nieuwerburgh, and Mindy Z. Xiaolan. 2019. "The U.S. Public Debt Valuation Puzzle." Working Paper 26583. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26583.

Jiang, Zhengyang, Hanno Lustig, Stijn Van Nieuwerburgh, and Mindy Z. Xiaolan. 2020. "Manufacturing Risk-Free Government Debt." Working Paper 27786. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27786.

Jiang, Zhengyang, Hanno Lustig, Stijn Van Nieuwerburgh, and Mindy Z. Xiaolan. 2021. "What Drives Variation in the U.S. Debt/Output Ratio? The Dogs That Didn't Bark." Working Paper 29351. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29351.

Joyce, Michael A. S., Ana Lasaosa, Ibrahim Stevens, and Matthew Tong. 2011. "The Financial Market Impact of Quantitative Easing in the United Kingdom." *International Journal of Central Banking* 7, no. 3: 113–61.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen. 2011. "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy." *Brookings Papers on Economic Activity*, Fall, 215–65.

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

Lucas, Deborah. 2012. "Valuation of Government Policies and Projects." *Annual Review of Financial Economics* 4:39–58.

Lustig, Hanno N., Stijn Van Nieuwerburgh, and Adrien Verdelhan. 2013. "The Wealth-Consumption Ratio." *Review of Asset Pricing Studies* 3, no. 1: 38–94.

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics* 15, no. 2: 145–61.

Mehrotra, Neil R., and Dmitriy Sergeyev. 2021. "Debt Sustainability in a Low Interest Rate World." *Journal of Monetary Economics* 124, suppl.: S1–S18.

Mian, Atif R., Ludwig Straub, and Amir Sufi. 2021. "A Goldilocks Theory of Fiscal Deficits." Working Paper 29707. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29707.

Portfolio Solutions Group. 2022. *Capital Market Assumptions for Major Asset Classes*. https://www.aqr.com/Insights/Research/Alternative-Thinking/2022-Capital-Market-Assumptions-for-Major-Asset-Classes.

Reis, Ricardo. 2021. "The Constraint on Public Debt When $r <$ g but $g <$ m." Working Paper 939. Basel: Bank for International Settlements. https://www.bis.org/publ/work939.htm.

Roll, Richard. 1977. "A Critique of the Asset Pricing Theory's Tests Part I: On Past and Potential Testability of the Theory." *Journal of Financial Economics* 4, no. 2: 129–76.

Santos, Manuel S., and Michael Woodford. 1997. "Rational Asset Pricing Bubbles." *Econometrica* 65, no. 1: 19–57.

Stambaugh, Robert F. 1982. "On the Exclusion of Assets from Tests of the Two-Parameter Model: A Sensitivity Analysis." *Journal of Financial Economics* 10, no. 3: 237–68.

Summers, Lawrence H. 2015. "Demand Side Secular Stagnation." *American Economic Review* 105, no. 5: 60–65.

van Binsbergen, Jules, Michael Brandt, and Ralph Koijen. 2012. "On the Timing and Pricing of Dividends." *American Economic Review* 102, no. 4: 1596–618.

van Wijnbergen, Sweder, Stan Olijslagers, and Nander de Vette. 2021. "Debt Sustainability When R – G < 0: No Free Lunch After All." Working Paper. Social Science Research Network, January 4. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3736921.

Comments and Discussion

COMMENT BY

WILLIAM GALE This paper applies modern finance techniques to analyze the federal budget outlook. The main conclusions are consistent with two long-held consensus findings that use more basic techniques. First, the nation has a long-term fiscal problem and will likely need to raise taxes or cut spending growth or both in the future (Auerbach 1994; CBO 2001). Second, higher interest rates make the government's fiscal situation substantially worse, both because the government is a net debtor (CBO 2022; Auerbach and Gale 2022) and—according to the authors—because of a maturity mismatch between government assets and liabilities.

OVERVIEW The main result is that the current market value of federal debt is larger than the present value of expected future primary surpluses, a condition that violates rational models. Proving this relation requires an estimate of the market value of debt, a projected path for future primary surpluses or deficits, and a discount rate.

Although most previous analyses of the fiscal outlook have used the par value of debt, the market value and par value of outstanding federal debt tend to track each other closely over time The market and par values of marketable Treasury debt in December 2021 were 104.7 percent and 101.0 percent of 2021 GDP respectively.[1]

To project budget outcomes, the authors use the Congressional Budget Office's (CBO) "current law" projections, which report primary deficits that average 3.2 percent of GDP over the next thirty years and rise as a share

1. Federal Reserve Bank of Dallas, "Market Value of U.S. Government Debt," https://www.dallasfed.org/research/econdata/govdebt#data.

Brookings Papers on Economic Activity, Fall 2022: 210–229

of GDP over time. As noted in the paper, the current law baseline is not a forecast of likely outcomes. Rather, it is the result of a variety of required assumptions, including the assumptions that there are virtually no changes in policy except reauthorization of spending programs, continued payment of full benefits in entitlement programs even if the trust funds are exhausted, discretionary spending rising with inflation rather than fixed in nominal terms, and increases in the debt ceiling to accommodate those changes (CBO 2022). For years after 2052, the authors assume the existence of persistent surpluses equal in present value to the value of the outstanding debt in 2052.

There is an extensive discussion of risk-adjusted discount rates. To bias the results against their main finding, the authors understate what they view as the appropriate market-based discount rate by assuming that tax revenues (which in practice are more pro-cyclical than the economy) and spending (which in practice is countercyclical) are both as risky as GDP itself.

In any case, the main point of the paper is that even with optimistic assumptions about future budget outcomes and conservative assumptions about discount rates, the market value of federal debt is (far) greater than the present value of expected future primary surpluses when discounted with risk-adjusted rates. Allowing for the government to collect resources via seigniorage does not change the basic conclusion.

POTENTIAL EXPLANATIONS The authors offer three potential explanations of the results: (1) Treasury market participants expect future fiscal corrections relative to the stated deficit path; (2) participants expect that fiscal policy will become pro-cyclical over time rather than remaining countercyclical; and (3) participants are mispricing Treasury debt. I discuss each of these in turn.

Given the fiscal outlook, the first explanation—that market participants expect fiscal corrections to be larger than posited by the authors—is plausible. This is standard fiscal reform: raise revenues or reduce spending relative to the baseline. It is worth noting that although the market values debt using a risky discount rate, the required fiscal changes to reach a given debt target in a given year are the same as in a non-stochastic framework. For example, the authors show in table 5 that an immediate and sustained 6.0 percent of GDP increase in the primary surplus would reduce the debt to –6.1 percent of GDP by 2052. Calculations using the non-stochastic model in Auerbach and Gale (2022) and the same budget projections generate the same answer. This is because the government's debt dynamics are governed by the rate of the interest the government pays even if investors are discounting the debt at a different, risky rate.

The second possibility—that the automatic stabilizer role of fiscal policy will be eliminated—seems the least plausible. Even if Congress wanted to do this, it would be difficult to implement, given that it would likely require a major redesign of core tax and spending programs. In principle, Congress could instead use discretionary tax and spending changes to more than offset the cyclical effects of the automatic stabilizers, but this seems unlikely. In addition, eliminating the countercyclical nature of fiscal policy would have severe consequences. Automatic stabilizers help stabilize the economy as a whole, and they provide critical assistance to people at precisely the time they need it (Edelberg, Sheiner, and Wessel 2022). It seems like we should be expanding automatic stabilizers; restricting them seems like a cure worse than the disease.

The authors call the third explanation "mispricing," but I would call it "discounting that is inconsistent with the model," to allow for the possibility that the discrepancy is due to model misspecification, not errors by market participants. I believe there is a plausible story for why market participants may use a lower discount rate for Treasuries than the authors propose.

Suppose that policymakers adjust primary surpluses in order to be sure to pay back the debt. That is, suppose primary surpluses are endogenous to debt issues (Auerbach and Gale 2009). Note that the current law baseline that the authors use assumes essentially no change in government behavior over the next thirty years and no changes in behavior in the case of unexpected shocks to the debt and thus misses this endogeneity. This seems to have several implications.

First, this policy endogeneity explains why the government can issue debt (at low rates) even if the present value of projected future primary surpluses is far less than current market value of debt: if investors know the government prioritizes debt repayment and avoidance of default, they will rationally expect future policies to adjust.

Second, it means that owning a government bond is different from owning a share of future primary surpluses that are projected assuming no change in policy (as in the current law baseline). When I buy debt, there is a promise that it will be paid back. That promise is embodied in the Fourteenth Amendment of the Constitution and is implicit in government actions. Despite massive fluctuations in primary deficits and surpluses, the government has prioritized paying back debt over other forms of spending since the War of 1812, with the exception of an administrative error in 1979 (Gale 2019).

Third, if policymakers prioritize debt repayment, it means that bondholders are not the residual claimants of risk, future taxpayers are. So, the paper may be showing that the risk-adjusted debt burden on future taxpayers is higher than commonly thought, not that government debt is riskier to market participants than commonly thought. If so, the idea that the correct risk-adjusted discount rate is lower than the authors assume seems plausible (Falkenheim 2021).

FISCAL CAPACITY So far, I have not yet mentioned the term "fiscal capacity," which the authors define differently from the rest of the literature. One would expect that having debt exceed something called "fiscal capacity" would be a bad thing, but that is not necessarily the case with the authors' definition. To be clear, using a different definition is perfectly fine—one can define terms however one would like; but different definitions have different implications, and the definition used by the authors does not have the implications that the authors seem to want to impose on it.

An intuitive definition of fiscal capacity would be the sum of current debt plus fiscal space, where fiscal space is the amount of new debt the government could add to its existing stock without adverse consequences. Those consequences could include, for example, disrupted financial markets, a recession, a default, or major capital outflows—in short, effects above and beyond the usual effects of debt.

These definitions are essentially what the OECD (Botev, Fournier, and Mourougane 2016) and IMF (Heller 2005) use and are grounded in actual, observable debt levels. Moreover, these definitions offer a useful guide to fiscal policy actions. In particular, knowing where the economy stands relative to these definitions would indicate whether the government could safely add new debt, whether issuance of new debt should face a higher bar than existing debt, and so on. In an earlier paper (Jiang and others 2021), the authors define fiscal capacity as "how much debt the government can issue," which I interpret as close to and consistent with this definition.

In this paper, however, the authors use a different definition. They define fiscal capacity as the present value of future primary surpluses (in some cases, adding in the seigniorage from the convenience yield on Treasury debt). Note that, unlike the OECD/IMF definition, this definition does not start with actual, observable debt levels; rather, it is based on infinite time horizon budget projections, which contain enormous amounts of uncertainty.

Second, importantly, this definition does not say that the government should stop issuing debt when it hits fiscal capacity. It does not say that there would be deleterious consequences of issuing debt above fiscal capacity—indeed, there may well be beneficial consequences of having debt exceed fiscal capacity. Suppose, for example, that the United States had zero debt and was expected to run balanced primary budgets starting now for eternity. Everyone would view that as a very strong fiscal position except the authors, who would say that fiscal capacity is zero (or even negative, given that the stream of tax revenues is riskier than the stream of outlays). But the United States clearly could issue new debt in that circumstance. More importantly, for many reasons (such as to provide the convenience that Treasury debt offers investors or to finance infrastructure, anti-recession policies, or wars), the government likely should issue substantial debt in that circumstance.

So it is unclear what the implications are for having the current market value of government debt exceed the authors' definition of fiscal capacity. To the extent that it means there is a chance we will need to raise taxes or reduce spending in the future, this observation is not wrong, but not new either.

CONCLUSION The paper is motivated by a desire to provide a summary assessment of the fiscal stance of the government. But the federal government is the most complicated financial institution in the world, and fiscal policy has many dimensions. For example, the issue is not just the deficit, how the money is raised and spent matters, too—in particular, deficits that finance investment in people or projects may have quite different effects than deficits that finance consumption (Gale 2019).

As a second example, the government could pay off its debt with future primary surpluses, as the paper notes, or with its stock of financial assets, which the paper ignores. Those assets equaled 7.3 percent of GDP at the end of 2021 (CBO 2022). Thus, the correct inequality would compare the present value of future primary surpluses with federal debt net of federal financial assets. The latter (in par value) equals 92.3 percent of GDP, so the authors' main results would still comfortably hold.

Recently, Blanchard (2019) has emphasized the importance of thinking about fiscal policy when interest rates are lower than the growth rate. Furman and Summers (2020) proposed that a useful criterion is to see whether real net interest payments are below 2 percent of GDP. The current paper argues that these criteria do not constitute sufficient statistics for assessing the entire fiscal situation. I agree completely, but I do not believe that the authors' measure of fiscal capacity is a sufficient guide to fiscal policy choices either. I would not want policymakers to base their choices solely on any individual criterion, but each criterion can be useful, helpful, and constructive.

REFERENCES FOR THE GALE COMMENT

Auerbach, Alan J. 1994. "The US Fiscal Problem: Where We Are, How We Got Here, and Where We're Going." *NBER Macroeconomics Annual* 9:141–86.

Auerbach, Alan J., and William G. Gale. 2009. "Activist Fiscal Policy to Stabilize Economic Activity." In *Economic Policy Symposium Proceedings: Financial Stability and Macroconomic Policy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Auerbach, Alan J., and William G. Gale. 2022. *The COVID Pandemic and the Federal Budget*. Washington: Brookings Institution. https://www.brookings.edu/research/the-covid-pandemic-and-the-federal-budget/.

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

Botev, Jarmila, Jean-Marc Fournier, and Anna Mourougane. 2016. "A Re-assessment of Fiscal Space in OECD Countries." Working Paper 1352. Paris: Organisation for Economic Cooperation and Development. https://www.oecd-ilibrary.org/economics/a-re-assessment-of-fiscal-space-in-oecd-countries_fec60e1b-en.

Congressional Budget Office (CBO). 2001. *The Budget and Economic Outlook: Fiscal Years 2002–2011*. https://www.cbo.gov/sites/default/files/107th-congress-2001-2002/reports/entire-report.pdf.

Congressional Budget Office (CBO). 2022. *The Budget and Economic Outlook: 2022 to 2032*. https://www.cbo.gov/publication/57950.

Edelberg, Wendy, Louise Sheiner, and David Wessel, eds. 2022. *Recession Remedies: Lessons Learned from the U.S. Economic Policy Response to COVID-19*. Washington: Brookings Institution. https://www.brookings.edu/wp-content/uploads/2022/04/RR-Complete-Volume.pdf.

Falkenheim, Michael. 2021. "Governmental Risk Taking under Market Imperfections." Working Paper 2021-07. Washington: Congressional Budget Office. https://www.cbo.gov/publication/57255.

Furman, Jason, and Lawrence Summers. 2020. "A Reconsideration of Fiscal Policy in the Era of Low Interest Rates." Discussion Paper. https://www.brookings.edu/wp-content/uploads/2020/11/furman-summers-fiscal-reconsideration-discussion-draft.pdf.

Gale, William G. 2019. *Fiscal Therapy: Curing America's Debt Addiction and Investing in the Future*. New York: Oxford University Press.

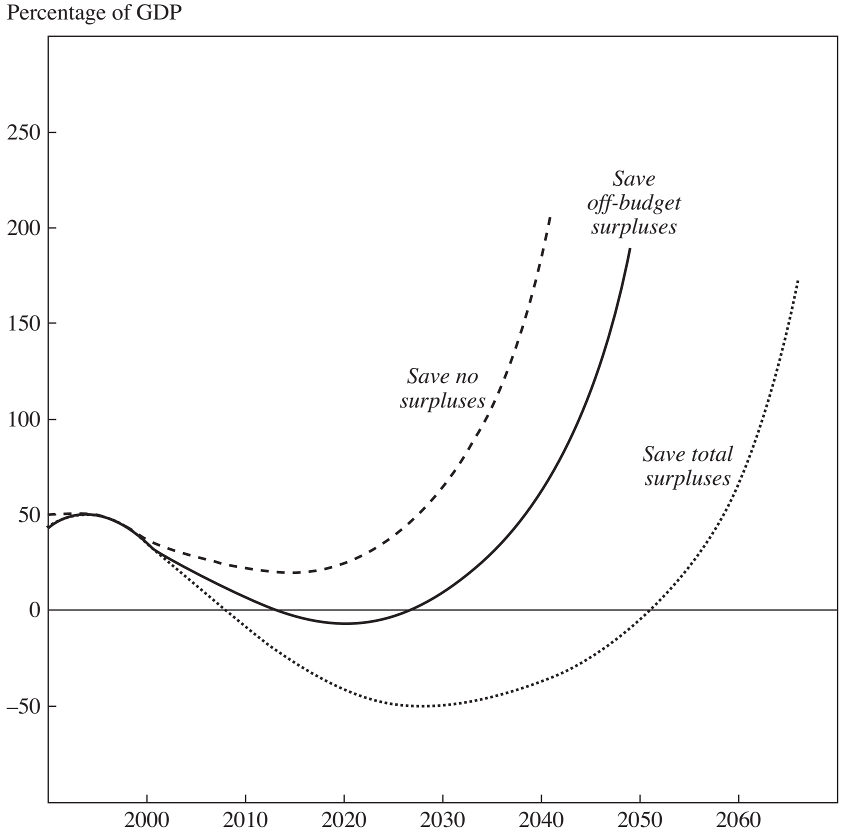
Heller, Peter. 2005. "Understanding Fiscal Space." Discussion Paper 2005/004. Washington: International Monetary Fund. https://www.imf.org/en/Publications/IMF-Policy-Discussion-Papers/Issues/2016/12/31/Understanding-Fiscal-Space-18065.

Jiang, Zhengyang, Hanno Lustig, Stijn Van Nieuwerburgh, and Mindy Z. Xiaolan. 2021. "The U.S. Public Debt Valuation Puzzle." Discussion Paper DP16082. London: Centre for Economic Policy Research.

COMMENT BY

DEBORAH LUCAS When I started my new job as chief economist at the Congressional Budget Office (CBO) twenty-two years ago, my first assignment from then CBO director Dan Crippen was to breathe some life

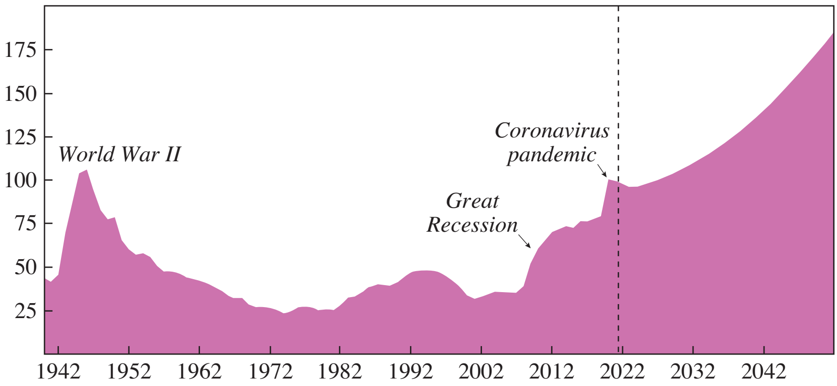
Figure 1. The CBO's Year 2000 Projections of Debt/GDP



Source: Reproduced from Congressional Budget Office (2000).

and urgency into the prose in the CBO's *Long-Term Budget Outlook*. The CBO's projections suggested that the current policy was unsustainable, but what could we say to get citizens and policymakers to pay attention? Most people, I'd venture even the CBO's staff, were largely indifferent. The situation was also a technical headache. When the CBO's equilibrium macro model was calibrated with its fiscal projections, there was nothing that we could do to prevent the model from crashing. Government debt crowded out private investment at an increasing rate over time, interest rates exploded, and the capital stock fell to zero—the real economy essentially vanished. Notably, the worst-case scenario in the CBO's 2000 *Long-Term Budget Outlook* (figure 1) understates the current debt-to-GDP ratio, primarily

Figure 2. The CBO's Year 2021 Projections of Debt/GDP



Source: Reproduced from Congressional Budget Office (2022).

because of elevated spending in response to the Great Recession and the COVID-19 pandemic.

Dire warnings about limits to fiscal capacity and the unsustainable path of projected fiscal policy long predated my initial forays into those issues at the CBO, and I expect they will become even more salient in the years and perhaps decades to come. Certainly, this year's edition of the CBO's long-term budget outlook is a close cousin of that inaugural edition more than two decades ago, and an even more pessimistic one (figure 2).

Assessments of the fiscal capacity of the US government seem to be something of a political Rorschach test. While many influential voices in academia, government, and the private sector continue to warn about the risks of excessive federal indebtedness, there are more than a few prominent economists who take a much more sanguine perspective. A tolerance for historically high debt ratios may arise from a sense of urgency for the federal government to confront pressing or even existential policy challenges today—tackling climate change, updating infrastructure, investments in social justice, and so forth. Recognizing that the only politically feasible way to fund such endeavors is with deficit spending, one might conclude that the best policy is to spend now and deal with the consequences later. There may be the expectation, or at least the hope, that future fiscal adjustments can accommodate the accumulated debt without too much pain, and that the public investments made today will make those high levels of debt more affordable in the future.

A relatively new twist to the question of whether and when deficit spending needs to be reined in, and one that has raised heated arguments on both sides, was ignited by the historically low interest rates of the last decade juxtaposed with high spending levels. I'll call that the $r - g$ debate. That's essentially where this paper comes in. The $r - g$ debate moved the conversation from the costs, benefits, and risks of high deficit spending to a seemingly technical set of issues concerning economic growth rates and appropriate discount rates. If the interest rate, r, on government debt is less than the growth of the economy, g, indefinitely, it is theoretically possible to grow out of high debt levels—ergo a much higher debt capacity.

The analysis here adds to a growing body of work that emphasizes that there is a fundamental problem with directly comparing r and g: the comparison isn't economically meaningful because it ignores the relation between risk and return. Specifically, treating growth opportunities as if their returns are risk-free but greater than the risk-free rate essentially assumes an arbitrage opportunity for the government. That observation is in the spirit of Barro (2020), my *BPEA* discussion earlier this year (Lucas 2021), Reis (2021), and other recent commentaries. A strength of this paper is to take seriously the importance of risk adjustment for fiscal policy evaluation. In that regard, it adds to the literature that estimates the market value of various fiscal obligations.[1]

DISCOUNTING FUTURE SURPLUSES The analysis of fiscal capacity rests on a derivation from the authors' earlier paper (Jiang and others 2019). There they showed that in the absence of bubbles and under rational expectations in a stationary stochastic economy, it is an identity that the value of Treasury debt at any point in time must equal the present value of future net government cash inflows excluding interest payments. In this paper, those net cash inflows are equated with primary surpluses. The authors then define fiscal capacity as the net present value of future primary surpluses.

The authors explore questions that include: How high must long-run primary surpluses be in order to cover the value of current Treasury debt liabilities? And how does risk adjustment affect one's conclusions about the answer? A robust result is that risk adjustment unambiguously reduces the present value of future primary surpluses when surpluses are proportional to GDP. Risk adjustment further reduces the present value of future surpluses when fiscal policy is countercyclical. Risk adjustment therefore suggests

1. To cite one closely related example, Geanakoplos and Zeldes (2010) contrast the market value of Social Security payment obligations, whose risk to the government is related to GDP, with the value calculated on an actuarial basis.

that larger future fiscal adjustments will be necessary than those implied by a similar analysis using Treasury rates for discounting.

Since risk adjustment is central to the analysis, I'll start with a quick reminder of why risk-adjusting discount rates is important, and why it shrinks the present value of primary surpluses. The fundamental economic reason to risk-adjust discount rates is that a unit of future consumption is worth more, in terms of today's consumption, when aggregate resources are scarce than when they're plentiful. That follows directly from decreasing marginal utility of consumption. It explains why the expected return on stocks is higher than on government debt, and why people are willing to pay more than an actuarily fair price for consumption insurance, effectively discounting its payoffs at less than a risk-free rate. The same logic applies when valuing risky fiscal cash flows.

The analysis in the paper rests on what-if experiments based on alternative calibrations of equation (1):

$$\text{(1)} \qquad \text{Debt}\big(\text{year-end 2021}\big) + \text{PV of future deficits}\big(\text{2022 to 2052}\big)$$
$$= \text{PV of primary surpluses}\big(\text{2052 to } \infty\big).$$

This equation is the identity referred to earlier, which was derived by iterating forward the government's flow budget constraint under the assumption of no bubbles in government debt prices (i.e., satisfying a transversality condition requiring the present value of time t debt to go to zero as t goes to infinity). Note that, like the authors, I've broken out the present value of deficits over the next thirty years from the total primary surplus stream to highlight that the CBO's projections are assumed to hold on average over the next thirty years and that positive surpluses will only start to be realized after that. I will note here that this is a strong assumption, and one I think would be violated if it appeared that confidence in US government debt was eroding.

Surpluses are expected to vary positively with GDP because tax collections are pro-cyclical and spending is countercyclical. The positive correlation with GDP implies a risk-adjusted rate that exceeds the risk-free rate. The simplest case is when surpluses are a constant share of GDP. A claim that is proportional to GDP is valued by discounting at a rate, r_{RA}, that includes a GDP risk-premium. If GDP and hence surpluses grow on average at a rate g, the present discounted value of the surpluses is approximately:

$$\text{(2)} \qquad \frac{\bar{s} \times \text{GDP}}{r_{RA} - g}.$$

The higher the risk-adjusted rate, the lower the present value of the future surpluses. The fact that fiscal policy is countercyclical means that surpluses have a "GDP beta" of greater than one, which would further increase r_{RA}. If the risk-adjusted rate is less than or equal to the growth rate of GDP, debt capacity is infinite.

Plugging in a range of plausible values for r_{RA} and g into equation (2) quickly reveals the enormous sensitivity of estimates of debt capacity that are based on discounted values over infinite horizons to assumptions about rates. While this sort of calculation is useful as a refutation of simple (risk-free) $r - g$ logic, the sensitivity to parameter choices suggests caution in using it to draw sharp conclusions about fiscal capacity.

The bottom line on the authors' choice for a baseline GDP risk premium is that it seems reasonable, and similar to what might be expected to emerge from other estimation approaches. The conclusions that the relevant risk-adjusted discount rate is effectively greater than the growth of GDP and that there is no free lunch in deficit spending when risk is accounted for are consistent with the authors' own previous work and with other analyses such as those mentioned earlier that have examined the implications of risk-adjustment for debt capacity.

Nevertheless, I will briefly quibble with the approach the authors took to identifying the risk-adjusted discount rate for GDP-linked claims. A technical concern is that the GDP risk premium is inferred with reference to unlevered stock returns. A large component of GDP is tied to labor income, which is very weakly correlated with stock market dividends. Dividends account for only a modest portion of capital income. There are two alternative approaches that would more directly link the premium to GDP risk and that would have been more convincing to me. The first would have been to use a utility-based macro model, for instance, like the one just rolled out by Newell, Pizer, and Prest (2022) for evaluating the discount rate for greenhouse gases. As well as ensuring that the risk premium was derived directly from the statistical properties of GDP, it would provide macroeconomists and policymakers with a more familiar point of reference. The second alternative, which could be used as a complement to the first, would be to estimate GDP betas for taxes and revenues using a model similar to the capital asset pricing model.

Beyond choosing a discount rate to apply to cash flows that are proportional to GDP, the authors had to make assumptions about risk associated with future surpluses and future debt levels. They make the important observation that US surpluses are pro-cyclical because spending rises and taxes fall during downturns. That provides valuable consumption insurance to citizens, and it is more costly to the government than a policy where surpluses are proportional to GDP. As the authors note, rich countries with high debt capacity reap considerable welfare benefits from the ability to run a countercyclical fiscal policy. A striking statistic from the International Monetary Fund is that advanced economies spent on average 11 percent of GDP on pandemic relief, whereas emerging markets spent about 4 percent. Preserving fiscal capacity is insurance that such policies will be feasible during the next crisis, and the ones after that.

I am less comfortable with some of the other assumptions that affect the fiscal capacity estimates, and those concerns are noted briefly here.

- The co-integration of debt with GDP is asserted without offering empirical support, and figure 2 suggests it may be counterfactual at least historically.
- It isn't clear to me why a focus on steady states is relevant during a period of unprecedentedly high peacetime debt ratios.
- CBO projections are not forecasts, and they are likely to deviate from expected outcomes, particularly over long horizons. In particular, they don't include legislative actions that reduce out-year deficits, even if such changes are viewed as likely.
- Related, my biggest concern is with the assumption that taxes and spending are on autopilot, whereas in fact policymakers are likely to adjust them in response to emerging stresses in the government debt market. Such adjustments could significantly reduce the risk of the debt and increase surpluses. That issue is further explored in the rest of this discussion.

IS CURRENT TREASURY DEBT OUTSTANDING RISKIER THAN INVESTORS THINK IT IS? The analysis raises the question, If the value of Treasury debt rests on risky primary surpluses in the distant future, can even very short-term Treasury debt rationally be considered by investors to be virtually risk-free? The paper addresses this issue only obliquely, and I found the discussions in the paper related to this issue quite confusing.

Equation (1) above implies that all Treasury debt is potentially risky, and its market price at any point in time should reflect that risk assuming rationality and no bubbles. It is important to emphasize this, as it was a point I didn't fully absorb until the authors pointed it out to me after my partially misleading remarks during the conference. The observations that follow have been revised to be consistent with the fact that the debt pricing model incorporates the possibility of default.

Despite all Treasury debt being risky because of its ultimate backing by uncertain future surpluses, I believe there is a strong case for investors to rationally believe that their current debt holdings are quite safe. Rather than estimating cash flows and evaluating default risk based on the unknowable distribution of possible paths for long-run primary surpluses, investors rationally expect to be paid in full as long as the government can garner the resources to make the promised payments and it has the legal authority to do so. It is reasonable to assume that those conditions will be satisfied in the near to medium term, as they almost always have been in the past. For longer-term Treasury debt, while there is clearly the possibility of a partial or even full default in some eventualities, it is reasonable for investors to expect losses to be small.

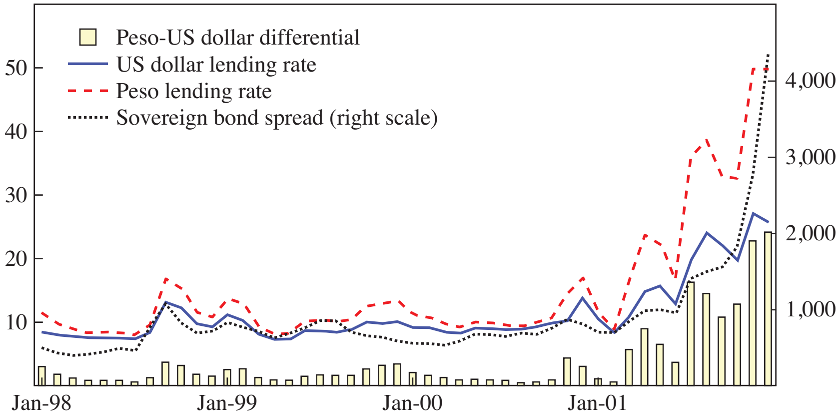
To put it differently, if the government treats public debt as a senior obligation, it will prioritize those payments over other types of spending. As for a firm, the seniority of debt makes it safer, while causing other claims to be riskier. My conjecture is that to reconcile the rationally perceived safety of the current debt with the identity in equation (1), it would require explicitly linking the surplus process to debt payment obligations. That would entail a surplus process that is different than the ones considered in this paper, but it need not violate the transversality condition. Surpluses could continue to be pro-cyclical in most but not all circumstances.

The fact that government debt can carry a very low interest rate even if fiscal capacity is quite limited has an important implication. Policymakers should not look to Treasury interest rates for reassurances that fiscal policies are sustainable or that they will be able to rely on debt-financed spending in the face of the next big crisis. The experience of less developed countries shows that government interest rates can shoot up very suddenly, as happened, for example, in Argentina in 2001 (figure 3). Note that equation (1) doesn't preclude a sharp and sudden downward revaluation of debt. It simply reflects that at any point in time, debt valuations will depend on current expectations about future fiscal policy and the economy. However, because the equation in itself is an accounting identity and not a stochastic model, it can't be used to predict whether and when stresses are likely to materialize.

SHOULD WE BE (MORE) WORRIED? The analysis suggests that the United States is out of fiscal capacity unless future surpluses are assumed to be implausibly large. However, there are a number of reasons to think that the government has more fiscal space than is suggested by these calculations.

My suspicion that fiscal capacity is underestimated is supported by consideration of the differences between reported primary surpluses and the actual resources the government has available to meet its debt obligations. The excess of actual resources over primary surpluses might be thought of as shadow surpluses.

Figure 3. Rapid Rise of Sovereign Debt Rates in Argentina



Source: Daseking and others (2005); reproduced with permission from International Monetary Fund.
Note: Rates are on thirty-day loans to prime customers.

Those shadow surpluses arise from the additional assets the government has available beyond its ability to tax its citizens or generate seigniorage, and from its ability to reduce spending if it is deemed necessary. Note that only capitalized tax revenues and seigniorage appear on the asset side of the federal balance sheet that is shown in the paper. The absence of nontax assets may reflect an implicit assumption that government expenditures are used for consumption rather than for investment.[2] A simple example illustrates how this can lead to an underestimate of government assets. Imagine that the government invests $1 billion in mortgage-backed securities in the open market, and that it funds that investment by issuing Treasury bonds. Under the budgetary rules governing asset purchases, the transaction increases the primary deficit by $1 billion. From an economic perspective the transactions are neutral; true fiscal capacity is unchanged but fiscal capacity as measured by the reported primary surplus falls. In fact, the largest (nontax) financial asset of the government is its $1.3 trillion student loan portfolio. While the market value is considerably less than the reported book value, its value is still substantial and it serves to offset a portion of the debt.[3]

2. Also missing on the liability side is an equity claim that is needed to absorb changes in the value of nontax assets.

3. Unlike for asset purchases, the budgetary accounting for student loans and other government credit programs is on an accrual basis. The use of accrual accounting for credit programs causes the deficit and the change in federal debt outstanding to diverge.

The government also has production technologies that it could use to increase revenues should the need arise. As the authors note, one such technology allows it to produce seigniorage, and its value is taken into account in some of the calculations. It can also produce citizenship rights that could be sold. It could sell public lands and increase prices on mineral and other natural resource rights. It could increase guarantee fees on the $5 trillion of mortgages it insures.

On the expenditure side, it has many levers to reduce costs or increase nontax revenues. For instance, it could increase co-payments in Medicare or end coverage of some expensive procedures. In the event of a war, it could cut military expenditures by reinstating the draft.

Even more dramatic actions could be taken. As the authors note, the government could rely on financial repression to force its citizens or domestic banks to hold its debt at below-market rates. It could take other actions to lower the value of outstanding debt, for instance, by expropriating foreign holders either directly or via a currency devaluation.

Perhaps most importantly, the Federal Reserve owns a large share of the debt held by the public, and it has the capacity to make additional very large purchases. It is difficult to predict what those purchases would imply for the value of the debt. However, the likelihood that the Federal Reserve would step in to prevent a default is a further reason why it is rational for investors to treat the promised payments as low risk in nominal terms.

A few caveats are in order. Many of these possibilities seem like very bad ideas. I also have provided no evidence that these adjustments would create significant additional debt capacity. However, they do suggest the possibility of a much higher debt capacity, and it would be interesting to explore their quantitative importance.

SHOULD THE GOVERNMENT INCREASE THE DURATION OF ITS DEBT? The authors emphasize that the duration of the government's debt liabilities is much shorter than the duration of its surplus assets. That duration mismatch causes fiscal capacity to be highly sensitive to interest rate risk. Lengthening the duration of its debt could reduce that risk, and this is suggested as a policy option.

The practicality of this advice is unclear. Issuing debt at anything close to the estimated duration of the surplus (283 years!) would create an asset that is incredibly risky for investors. Enticing investors to buy it presumably would require paying a substantial term premium. Effectively, the government would be buying insurance against interest rate risk from the private sector, an arrangement that would run counter to the usual presumption that the government has the greater risk-bearing capacity. Issuing very long-term nominal debt would also create a moral hazard problem because of the temptation to inflate away its value. Recognition of that risk could further increase the interest rate demanded by investors. Issuing long-term real debt would avoid the moral hazard problem, but I expect it would lack liquidity and also carry a hefty term premium. The Treasury chooses the maturity structure of the debt so as to minimize long-run funding costs and take into account factors like rollover risk. Adding a surplus hedging objective would complicate what is already a difficult optimization problem.

REFERENCES FOR THE LUCAS COMMENT

Barro, Robert J. 2020. "*r* Minus *g*." Working Paper 28002. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28002.

Congressional Budget Office (CBO). 2000. *The Long-Term Budget Outlook*. https://www.cbo.gov/publication/12749.

Congressional Budget Office (CBO). 2022. *The 2022 Long-Term Budget Outlook*. https://www.cbo.gov/publication/57971#section1.

Daseking, Christina, Atish R. Ghosh, Timothy D. Lane, and Alun H. Thomas. 2005. "Lessons from the Crisis in Argentina." Washington: International Monetary Fund. https://www.imf.org/en/Publications/Occasional-Papers/Issues/2016/12/30/Lessons-from-the-Crisis-in-Argentina-17402.

Geanakoplos, John, and Stephen P. Zeldes. 2010. "Market Valuation of Accrued Social Security Benefits." In *Measuring and Managing Federal Financial Risk*, edited by Deborah Lucas. Chicago: University of Chicago Press.

Jiang, Zhengyang, Hanno Lustig, Stijn Van Nieuwerburgh, and Mindy Z. Xiaolan. 2019. "The U.S. Public Debt Valuation Puzzle." Working Paper 26583. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26583.

Lucas, Deborah. 2021. Comment on "The Sustainability of State and Local Pensions: A Public Finance Approach," by Jamie Lenney, Byron Lutz, Finn Schüle, and Louise Sheiner. *Brookings Papers on Economic Activity*, Spring, 49–57.

Newell, Richard G., William A. Pizer, and Brian C. Prest. 2022. "A Discounting Rule for the Social Cost of Carbon." *Journal of the Association of Environmental and Resource Economists* 9, no. 5: 1017–46.

Reis, Ricardo. 2021. "The Constraint on Public Debt When $r < g$ but $g < m$." Working Paper 939. Basel: Bank for International Settlements. https://www.bis.org/publ/work939.htm.

GENERAL DISCUSSION Michael Falkenheim argued that a risk-adjusted projection of debt would have been more useful than a model focusing on current capacity; asking what the debt level in the future would have to be, in present value terms, to cover current debt and surpluses between now and 2052. That could have served as an alternative debt projection to the one produced by the Congressional Budget Office (CBO) and would better reflect the Arrow-Debreu considerations brought up by discussant Deborah Lucas. He continued, pointing out that in thinking about the capacity for debt, one may want to start by thinking about the capacity for a primary surplus in terms of economic and political sustainability. Finally, Falkenheim mentioned how, in an overlapping generations model such as that by Peter Diamond, there is an opportunity for the government to enact Pareto improving policies which may be considered a form of arbitrage—in which case the no-arbitrage condition would be violated.[1]

Ricardo Reis suggested that in addition to focusing on GDP risk, the authors may also want to include inflation risk in the model. In light of the inflation risk premium having been essentially zero—or even negative—over the past decade but rising over the last year, one could use the authors' calculations to find the cost of inflation in shrinking capacity through making the nominal debt riskier and having a higher interest rate. Connecting this to the paper by Ball, Leigh, and Mishra, in which a projection of two years of elevated inflation is put forward, Reis argued that using the authors' methods one could analyze how costly the resulting increase in the inflation risk premium would be fiscally.[2]

Henry Aaron commented that while he does not disagree with the notion of needing to narrow the gap between spending and revenues collected, he rejects the use of a current law assumption by the CBO in their projections. He pointed to Medicare hospital insurance and Social Security, where the CBO relies on the statutorily committed levels of spending thirty years into the future, even as the funds are exhausted—which they will be within a few years and about a decade, respectively. There are specific corrections that could be made, Aaron continued, which could bring the CBO numbers closer to what they claim to be—current law—but the more basic point is that the CBO's claim that its current long-term projections are based on current law is false and conceals what CBO really does. The CBO assumes that unreduced social insurance pension and health benefits will be paid even when trust funds are depleted, a policy that Congress has explicitly barred. And it assumes that Congress will cut income tax rates or other taxes to hold constant the tax/GDP ratio even when it projects that

1. Peter A. Diamond, "National Debt in a Neoclassical Growth Model," *American Economic Review* 55, no. 5 (1965): 1126–50.

2. Laurence Ball, Daniel Leigh, and Prachi Mishra, "Understanding US Inflation during the COVID-19 Era," in the present volume of *Brookings Papers on Economic Activity*.

the budget will be hugely in deficit, something that violates established congressional policies for normal times. These violations of the "we base our projections on current law" assumption have important economic and political ramifications.

In response to Aaron's comment Phillip Swagel noted that the CBO can provide the authors with the data Aaron mentioned. Swagel explained that the reason they adhere to the current statutorily committed levels of spending is because the CBO does not engage in predicting what a future Congress will do but, he explained, they do analyze alternative scenarios under which spending is more or less than the current statutorily committed levels.

Louise Sheiner commented that the CBO projections show what expected taxes would have to be for the government to meet its debt obligation under the assumption that this will ultimately fall on the taxpayer. She pointed out that this puts taxpayer risk rather than debt-holder risk in focus. This in turn makes business cycle risk less of a concern but does not eliminate GDP risk and potentially slow growth as important factors. For example, in a world of slow technological progress and low GDP growth where people do not live much longer as a result, Medicare spending may be quite a bit lower, she suggested, and in that case, when considering long-term productivity, some of the government spending may be offset.

Jonathan Parker suggested that one of the fundamental future risks is whether we will see a return of high trend output growth, which comes with higher real interest rates, or the reverse—sluggish productivity growth but low interest rates, which would make the debt easier to roll over but more of a long-term concern. This long-term risk in the growth rate should enter the analysis, he argued. Parker made a second point that there is a government budget constraint that must hold, and he noted that if there is a mis-valuation or lack of sustainability in the authors' analysis, then something in the future would have to fill that hole and that might be inflation, as it has been in many countries in the past.

Donald Kohn responded to the panelists' claim that the market must embody either a large fiscal correction sometime relatively soon or a lot more repression, saying he did not find either persuasive. Kohn pointed out that market expectations on inflation are low—close to the 2 percent target in the long-term. Perhaps there is a shorter horizon than the infinite horizon suggested by the authors or, he asked, are they thinking in terms of debt capacity more in the way the discussants were? Kohn concluded by asking for more discussion on the pattern of market prices.

Steven Davis wondered how the rest of the world fits into the authors' analytical framework. As emphasized by Lucas, the rest of the world holds a lot of US Treasuries, with an associated convenience yield. Davis contemplated a scenario in which the world grows rapidly relative to the United States and demand for US Treasury securities increases as a result—their role as reserves would be affected, effectively introducing risk. He then pointed to the possible emergence of good substitutes for the US dollar, creating more competition and potentially eroding the convenience yield altogether, suggesting that these considerations should have been factored into the authors' model.

Arvind Krishnamurthy reflected on the focus of the paper as not so much trying to establish fiscal capacity but rather providing input that can help us get to the fiscal capacity. He noted that the paper engages in a sort of valuation exercise, asking, What must you believe about the surplus process, and the interest rates applied to that process, in order to reconcile how much investors are willing to pay for US debt? Krishnamurthy pondered the different ways the authors go about reconciling this: to reconcile the government debt with just movements in the risk premium, the latter would have to be incredibly low. He then contemplated how increasing convenience yields may resolve the analysis, referring to the points made by Davis. Finally, Krishnamurthy addressed some of the comments on including inflation in the framework and suggested that the authors may have to relax their rationality assumption for this purpose, noting that if investors were expecting inflation, interest rates would have already adjusted. Therefore, in the authors' framework the assumption would have to be a world in which investors irrationally expect no inflation—which given the paper by Ball, Leigh, and Mishra may not be too far-fetched—and unanticipated inflation could then potentially help reconcile the valuation.

Jonathan Pingle suggested that considering market imperfections to a greater extent and how to reconcile those may be key to understanding the gap between the market pricing and what the optimal trajectory of capacity may be. Pingle argued that the convenience yield, in a sense, is a form of market imperfection. He noted that we tend to think about both Federal Reserve holdings of Treasuries and central bank holdings as having a different rollover risk than the private market and that research shows that central bank holdings put downward pressure on yields. He added that another important issue in the United States is money market reform, shifting the industry to be almost fully government-only funds following the Dodd-Frank reform. Pingle argued that while this won't be reflected in the convenience yield further out on the curve, there is a significant amount of issuance now that faces very little rollover risk and an additional type of market imperfection creating demand for federal government debt.

Justin Wolfers remarked that while the authors suggest that bonds are mispriced, whether this is correct or not does not necessarily affect anything beyond Wall Street. He argued that on Main Street, the average worker will still show up at work every day, regardless of the success of the valuation exercise by the authors.

Hanno Lustig clarified that the analysis indeed implies that debt cannot be risk-free—the debt could only be risk-free if the tax claim is less risky than the spending claim.

Stijn Van Nieuwerburgh commented that the exercise in the analysis was to apply aggregate risk pricing to the issue of fiscal sustainability. Van Nieuwerburgh summarized the feedback as focusing importantly on the gap between the present value of government surplus and the market value of government debt and noted first the issue of the convenience yield and the extent to which demand for US debt may rise in the future. He argued that given a downward sloping demand curve, a higher future convenience yield seemed unlikely.

To that point, Lustig noted that while the United Kingdom was once the world's safe asset supplier, earning large convenience yields, this ended abruptly with World War I after which the UK government had to borrow at a much higher interest rate. He argued that it would be foolish to assume that an alternative to the United States as a safe asset supplier would never exist.

Van Nieuwerburgh mentioned that another option that had been brought up was unexpected inflation and the possibility that bond investors were misunderstanding the inflation risk. He pointed out that bond investors may be systematically overpredicting the surplus, as they are overly optimistic about fiscal rectitude. An additional possibility is the presence of a bubble, which Van Nieuwerburgh found implausible. Finally, fiscal adjustment would be an option that, while politically undesirable, may become necessary in the future.

Zhengyang Jiang responded to the comments on pricing other risks in the model, including inflation and interest rate risk, and noted that the resulting net present value of surpluses comes out even lower in models where these other factors are included—even becoming negative. Thus, the question of what would close the gap remains, but presumably lies in some of the possibilities listed by Van Nieuwerburgh.

PANEL on SHRINKING THE FEDERAL RESERVE BALANCE SHEET

ARVIND KRISHNAMURTHY
Stanford University Graduate School of Business

Lessons for Policy from Research

ABSTRACT I review lessons from the research on central bank actions over the last decade and draw out implications for expanding the Federal Reserve balance sheet (quantitative easing) and shrinking the balance sheet (quantitative tightening). As I outline, there is already enough evidence in the research to indicate the manner in which the Federal Reserve could update its policy normalization principles and plans.

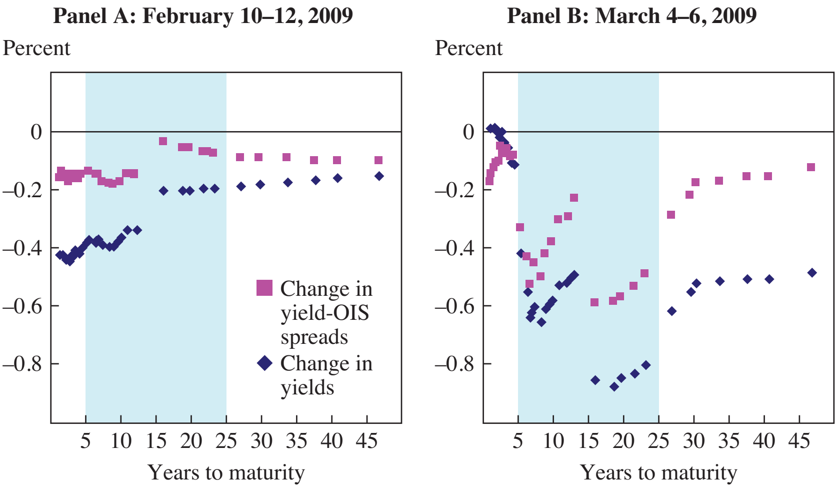
Former Federal Reserve chairman Ben Bernanke famously quipped, in a 2014 discussion at the Brookings Institution, that "the problem with QE is that it works in practice, but it doesn't work in theory." Academic and policy research on quantitative easing (QE) has come quite far over the last decade, and we are less in the dark about the workings of QE. In this paper, I review the lessons from this research and then draw out implications for expanding the Federal Reserve balance sheet (QE) and shrinking the balance sheet (quantitative tightening, or QT).

There are three principal lessons from the research: (1) QE works differently than conventional monetary policy in that the impacts are highest in the asset market targeted. (2) QE impacts are highest during periods of financial distress, market segmentation, and illiquidity. While this statement is likely also true of conventional policy, the effects are much more dramatic with QE. (3) QE alters the quantity of central bank reserves, and the post-2008 regulatory and economic regime implies substantially higher necessary reserve balances. I review each of these points and then turn to their implications for the formulation of rules governing QE/QT. The Fed

Conflict of Interest Disclosure: Arvind Krishnamurthy is a consultant to the Federal Reserve Bank of San Francisco. Other than the aforementioned, 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, Fall 2022: 233–242

Figure 1. Yield Changes by Maturity from UK QE for UK Gilts and Gilt-OIS Spreads



Source: Joyce and others (2011); copyright Bank of England and the Association of the International Journal of Central Banking; adapted with permission.

currently uses QE in two ways: to provide liquidity to markets during financial illiquidity episodes ("crisis QE") and to lower financing costs for borrowers at a time when the zero lower bound binds ("easing QE"). I argue that rules for these two types of policies should differ, but that the Fed has blurred the lines between them which has led to policy errors.

I. Lessons from Research

I.A. QE Works through Narrow Channels

Joyce and others (2011) present data from an event study around two significant QE news dates in 2009 by the Bank of England. On February 11, 2009, the *Inflation Report* and the subsequent press conference gave a strong indication that the bank would do QE. Markets interpreted this to mean that the bank would purchase bonds out to around fifteen-year maturity. On March 5, 2009, the bank announced that purchases would be in the five- to twenty-five-year range. Figure 1, replicating figure 4 in Joyce and others (2011), shows the changes in gilt yields around the event dates and the changes in the spread between gilt and overnight index swap (OIS) yields around these dates. Panel A shows the market reaction to the

February announcements: yields fall across the board. The pattern is similar to a conventional policy response in that there are larger effects on short-term bonds than longer-term bonds. In the curve showing the yield-OIS spread change, we see unique QE effects. If the policy transmission is akin to conventional monetary policy, there should be no change in these spreads as we would expect that both gilt yields and OIS yields will move in lockstep so that their spread would not change. Panel B shows the market reaction to the March announcement, and here we can really see the unique QE effects. First note that the effect on gilt yields is concentrated in the five- to twenty-five-year range, which the bank indicated as the target of QE purchases, with yields in the fifteen- to twenty-five-year range falling dramatically on the news that these maturities would also be purchased. Second, note that the yield-OIS spread change reflects the same pattern and similar magnitude, indicating that it is particularly gilt yields that are being affected by the announcement.

Following earlier work with Annette Vissing-Jorgensen (Krishnamurthy and Vissing-Jorgensen 2013), I refer to these effects as via a “narrow” channel rather than the “broad” channel of conventional monetary policy. That is, QE policy most affects the prices of the asset targeted in QE (gilts in this case). In contrast, conventional policy moves all asset prices from gilts to OIS rates and even stock prices. We offer much more evidence of these types of effects in examining the response of asset prices to news regarding the Fed’s QE purchases of mortgage-backed securities (MBS). They show narrow effects: the prices of the current coupon MBS, which is the asset purchased by the Fed, move the most relative to other coupon MBS, older MBS, and non-MBS assets (Krishnamurthy and Vissing-Jorgensen 2013).

There is more narrow channel evidence on the impact of QE. Eser and Schwaab (2016) show that the security markets program of the European Central Bank (ECB) lowered the targeted countries’ sovereign bond yields, particularly relative to non-targeted bonds. Grosse-Rueschkamp, Steffen, and Streitz (2019) and Todorov (2020) show that the ECB corporate sector purchase program lowered the bond yields of the eligible corporate bonds, particularly compared to non-eligible bonds’ yields. Moussawi (2022) shows a similar effect for the Fed’s municipal bond liquidity facility introduced during the COVID-19 crisis in 2020. Barbon and Gianinazzi (2019) likewise show that the Bank of Japan’s exchange-traded funds (ETF) purchase program affected eligible stock prices significantly relative to non-eligible ones.

The evidence I have cited concerns the impact of QE on asset prices. Of course it is more important to understand the effect of these asset price changes on decisions of economic agents. Here again, the evidence is most consistent with a narrow channel. Rodnyansky and Darmouni (2017) examine changes in bank lending, contrasting behavior across Fed purchases of MBS and purchases of Treasuries. They show that banks with significant holdings of MBS expand real estate lending after the MBS purchases, but not the Treasury purchases. This evidence is most consistent with a narrow channel impact of QE. Di Maggio, Kermani, and Palmer (2020) show that the Fed's MBS purchases particularly spur conforming, as opposed to jumbo, mortgage originations. This is narrow channel evidence because the Fed purchased conforming mortgages. There is analogous evidence from the behavior of firms in response to ECB corporate bond purchases. Grosse-Rueschkamp, Steffen, and Streitz (2019) show that the eligible firms in the ECB purchase program respond by issuing more bonds and borrowing less from banks, compared to non-eligible firms.

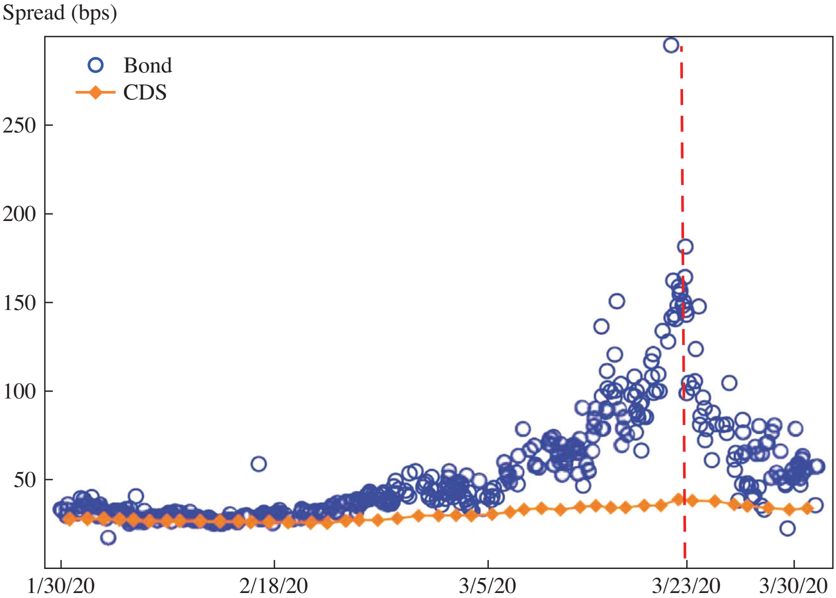
I.B. QE Impacts Are Highest during Periods of Financial Distress

Figure 2 plots the yield spread on Google's investment-grade six-year bond and the five-year credit default swap (CDS) for Google. The figure replicates figure 1 of Haddad, Moreira, and Muir (2021). We see the dramatic rise in the bond yield, reflecting the financial market dislocation at the start of the COVID-19 recession. The CDS rate does not change, reflecting that there is little change in the fundamental default risk of Google. The Fed announced the introduction of its corporate bond facilities on March 23, 2020, which allowed for purchases of investment-grade bonds. The yield spread declined dramatically with this announcement. It should be apparent that a similar announcement of a corporate bond facility say on February 1, 2020, would have had a very small effect on spreads. That is, the evidence here shows that QE impacts are highest during periods of financial distress. There is similar evidence by Gilchrist and others (2021) studying the Fed's corporate bond facility. Gorodnichenko and Ray (2017) provide related evidence for a longer sample showing that demand shocks in the market for US Treasury bonds have a much larger impact during periods of financial turmoil than during calm periods.

I.C. The Federal Reserve Balance Sheet Needs to Be Larger in 2022 than in 2008

Another important finding from research is that the minimum level of reserve balances needed to ensure a smooth functioning of the interbank

Figure 2. Google Bond Yield Spread and CDS



Source: Haddad, Moreira, and Muir (2021); adapted with permission from Oxford University Press and the Society for Financial Studies.

Note: Vertical dashed line indicates Federal Reserve corporate bond purchase program, announced March 23, 2020.

market is in excess of \$1.5 trillion, and considerably larger than the pre-crisis reserve balance of around \$60 billion. This is the finding of Copeland, Duffie, and Yang (2021), who examine the repo market dislocation in September 2019, concluding that the level of reserves at the time of \$1.4 trillion was too small given the regulatory and economic regime after the 2008 crisis. Afonso and others (2022), as well as Lopez-Salido and Vissing-Jorgensen (2022), estimate the banking sector's reserve demand function over the 2010s. Afonso and others (2022) show that reserve demand flattens at quantities of reserves of around 13 percent of bank assets, or in excess of \$2 trillion currently. Lopez-Salido and Vissing-Jorgensen (2022) predict that reserve demand flattens at quantities around \$3.5 trillion. While these numbers differ across research papers, they all indicate a substantially higher minimum reserve balance, running into the trillions of dollars, and hence a larger Fed balance sheet than the pre-2008 balance sheet, when reserves were on the order of tens of billions of dollars.

II. Lessons for Policy

Conceptually, there are two different types of QE the Fed has pursued. It has done crisis QE to alleviate systemic risk in an illiquidity episode, such as the COVID-19 corporate bond actions, and easing QE to reduce long-term rates, as the Fed did with its MBS purchases in 2010 to 2013 and again in 2020 to 2022. The research indicates that these two types of QE work differently—that the impacts of crisis QE are much larger than easing QE (point 2 above) and that easing QE has its largest impact on the asset market targeted (point 1 above).

The two types of QE suggest that the Fed should have two different rules governing QE/QT, but in practice, the Fed has followed a single rule. As I argue next, this has led to policy errors. Of most significance to the current 2022 tightening cycle, the Fed may have needlessly contributed to a housing market bubble that will now need to be popped.

II.A. Channel Fallacy

During the COVID-19 illiquidity period of March and April 2020 and during the recovery from the COVID-19 recession, the Fed purchased MBS. As noted in points (1) and (2) above, these purchases have their largest impacts during an illiquidity period and in the asset market targeted. That is, QE works through a narrow channel. The narrow effects meant that the MBS purchases (crisis QE) in the spring of 2020 were beneficial given the systemic liquidity stresses in the fixed-income market (Chen and others 2021). However, the Fed continued the MBS purchase program well after the period of liquidity stress ended, through 2020, and only ceased purchases and reinvestments in September 2022. This is a policy error that stems from not recognizing that MBS purchases work through a narrow channel and not the broad channel of conventional policy.

The MBS purchases outside an illiquidity episode are easing QE. These purchases brought down mortgage rates and had beneficial impacts in the recovery from the 2008 financial crisis because housing and mortgages were central to macro dynamics during that recovery. In 2021 and 2022, the MBS purchases to reduce mortgage rates can only be rationalized if the support to the housing market would have had beneficial spillovers to the rest of the economy. But there has been no evidence for that. Instead, the Fed's purchases may have needlessly contributed to a housing market bubble. As policy has shifted to a tightening mode, this housing boom now looms as a risk to financial stability.

II.B. The Ratchet Problem

The Fed's policy rule governing QE/QT is best described by what I call a "tying together" rule. The Fed expands the balance sheet in crisis states. But then it ties balance reductions to changes in the stance of conventional policy. In the Fed's 2014 statement on policy normalization, it outlined a plan to gradually raise its target range for the federal funds rate to more normal levels and gradually reduce the Fed's securities holdings to normal levels.[1] While these plans have been updated several times since 2014, the underlying approach to tie together balance sheet policy during QT with policy rate increases has remained.

Let us next consider what a QE/QT policy rule would look like in light of the research I have reviewed. If we index financial stress by x, then, the Fed should expand the balance sheet in states worse than x and shrink the balance sheet in states better than x. Here, x is determined by the cost of balance sheet size and the macro benefit of policy. Point (3) above—that the economy requires much higher reserve balances than was the case before the global financial crisis of 2008—is relevant to the cost of balance sheet size and the determination of x. Moreover, in states better than x, the Fed should shift from crisis QE to a smaller balance sheet size governed by the benefit of easing QE. The balance sheet should be smaller because the benefit of QE is smaller in normal states compared to crisis states.

In contrast to this policy rule, the Fed's tying-together policy has led it to delay balance reduction. The Fed has created a balance sheet ratchet, which the banking sector then adapts to, making it costlier to subsequently reduce the balance sheet, as argued by Acharya and Rajan (2022).

II.C. Communication

The tying-together rule also creates communications challenges for the Fed. A QT which is about a winding down of crisis QE and a QT which is about a winding down of easing QE send very different signals to economic agents. Suppose that investors see a QT which the Fed intends as the end of crisis QE but which investors misinterpret to be the end of easing QE. In this case, investors will then expect that the QT will be followed by increases in the policy rate, and this expectation will lead to an unintended

1. Board of Governors of the Federal Reserve System, "Policy Normalization," https://www.federalreserve.gov/monetarypolicy/policy-normalization-discussions-communications-history.htm.

tightening of monetary conditions for conventional reasons. If, on the other hand, the Fed was able to communicate its intent clearly, then this effect would not arise, and indeed agents may see the end of crisis QE as good news regarding the health of the financial system.

The best example of this communication breakdown is the taper tantrum of 2013. In Krishnamurthy and Vissing-Jorgensen (2013), a paper prepared for the Jackson Hole Symposium, we argued that the taper tantrum occurred because the Fed communicated that it would undertake QT, which then led the market to conclude that the Fed would also raise the policy rate. In short, the market anticipated that if QE was no longer required, then the zero lower bound would no longer be a constraint on policy.

III. Conclusion

Research over the last decade has shed considerable light on the ways in which QE works. I have outlined how this research can inform the rules governing QE/QT. While more research is needed on the workings of QE, there is already enough evidence in the research to indicate the manner in which the Fed could update its policy normalization principles and plans. Not doing so will likely lead to more errors of the kind that I have described.

ACKNOWLEDGMENTS I thank Janice Eberly, Deborah Lucas, and Hanno Lustig for helpful conversations and Hala Moussawi for research assistance.

References

Acharya, Viral V., and Raghuram Rajan. 2022. "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. https://www.nber.org/papers/w29680.

Afonso, Gara, Domenico Giannone, Gabriele La Spada, and John Williams. 2022. *Scarce, Abundant, or Ample? A Time-Varying Model of the Reserve Demand Curve*. Staff Report 1019. Federal Reserve Bank of New York.

Bank of England. 2009. *Inflation Report—February 2009*. https://www.bankofengland.co.uk/inflation-report/2009/february-2009.

Barbon, Andrea, and Virginia Gianinazzi. 2019. "Quantitative Easing and Equity Prices: Evidence from the ETF Program of the Bank of Japan." *Review of Asset Pricing Studies* 9, no. 2: 210–55.

Bernanke, Ben. 2014. "Central Banking after the Great Recession: Lessons Learned and Challenges Ahead." Excerpt from a discussion with Federal Reserve Chairman Ben Bernanke on the Fed's 100th anniversary, Brookings Institution, Washington, DC, January 16. https://www.brookings.edu/wp-content/uploads/2014/01/20140116_bernanke_remarks_transcript.pdf.

Chen, Jiakai, Haoyang Liu, Asani Sarkar, and Zhaogang Song. 2021. *Dealers and the Dealer of Last Resort: Evidence from MBS Markets in the COVID-19 Crisis*. Staff Report 933. Federal Reserve Bank of New York.

Copeland, Adam, Darrell Duffie, and Yilin Yang. 2021. "Reserves Were Not So Ample After All." Working Paper 29090. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29090.

Di Maggio, Marco, Amir Kermani, and Christopher J. Palmer. 2020. "How Quantitative Easing Works: Evidence on the Refinancing Channel." *Review of Economic Studies* 87, no. 3: 1498–528.

Eser, Fabian, and Bernd Schwaab. 2016. "Evaluating the Impact of Unconventional Monetary Policy Measures: Empirical Evidence from the ECB's Securities Markets Programme." *Journal of Financial Economics* 119, no. 1: 147–67.

Gilchrist, Simon, Bin Wei, Vivian Z. Yue, and Egon Zakrajšek. 2021. "The Fed Takes on Corporate Credit Risk: An Analysis of the Efficacy of the SMCCF." Working Paper 963. Basel: Bank for International Settlements. https://www.bis.org/publ/work963.htm.

Gorodnichenko, Yuriy, and Walker Ray. 2017. "The Effects of Quantitative Easing: Taking a Cue from Treasury Auctions." Working Paper 24122. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24122.

Grosse-Rueschkamp, Benjamin, Sascha Steffen, and Daniel Streitz. 2019. A capital structure channel of monetary policy. *Journal of Financial Economics* 133, no. 2: 357–78.

Haddad, Valentin, Alan Moreira, and Tyler Muir. 2021. "When Selling Becomes Viral: Disruptions in Debt Markets in the COVID-19 Crisis and the Fed's Response." *Review of Financial Studies* 34, no. 11: 5309–51.

Joyce, Michael A. S., Ana Lasaosa, Ibrahim Stevens, and Matthew Tong. 2011. "The Financial Market Impact of Quantitative Easing in the United Kingdom." *International Journal of Central Banking* 7, no. 3: 113–61.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen. 2013. "The Ins and Outs of LSAPs." In *Economic Policy Symposium Proceedings: Global Dimensions of Unconventional Monetary Policy*. Jackson Hole, Wyo: Federal Reserve Bank of Kansas City.

Lopez-Salido, David, and Annette Vissing-Jorgensen. 2022. "Reserve Demand and Balance-Sheet Runoff." Working Paper. Washington: Board of Governors of the Federal Reserve System.

Moussawi, Hala. 2022. "Asset Purchases and Preferred Habitats in the Municipal Bond Market." Working Paper. Social Science Research Network, March 30. https://ssrn.com/abstract=4017698.

Rodnyansky, Alexander, and Olivier M. Darmouni. 2017. "The Effects of Quantitative Easing on Bank Lending Behavior." *Review of Financial Studies* 30, no. 11: 3858–87.

Todorov, Karamfil. 2020. "Quantify the Quantitative Easing: Impact on Bonds and Corporate Debt Issuance." *Journal of Financial Economics* 135, no. 2: 340–58.

SYDNEY C. LUDVIGSON
New York University

Market Reactions to the Federal Reserve's Balance Sheet Normalization Plans

ABSTRACT This paper focuses on interpreting the stock market's reactions to Federal Reserve announcements about its balance sheet normalization plans, applying the methodology developed with Francesco Bianchi and Sai Ma. The results indicate that the stock market declines after announcements, suggesting perceived inflexibility in statements about balance sheet normalization, but many of the large reactions to these announcements can be ascribed to forces that move the stock market but not the broader economy.

In this paper, I focus on interpreting the stock market's reactions to Federal Reserve announcements about its balance sheet normalization plans. To do so, I apply the methodology in recent work with Francesco Bianchi and Sai Ma that integrates a high-frequency monetary event study into a mixed-frequency macro-finance model and structural estimation (Bianchi, Ludvigson, and Ma 2022).[1] We begin with an event study of the major Fed communications pertaining to its balance sheet normalization plans.

1. The underlying code relies on the working paper. The replication materials will be made available when that paper is published.

Conflict of Interest Disclosure: Sydney C. Ludvigson's work was supported in part by the National Science Foundation (grant no. 2116641). Other than the aforementioned, 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, Fall 2022: 243–258 © 2023 The Brookings Institution.

I. High-Frequency Event Study

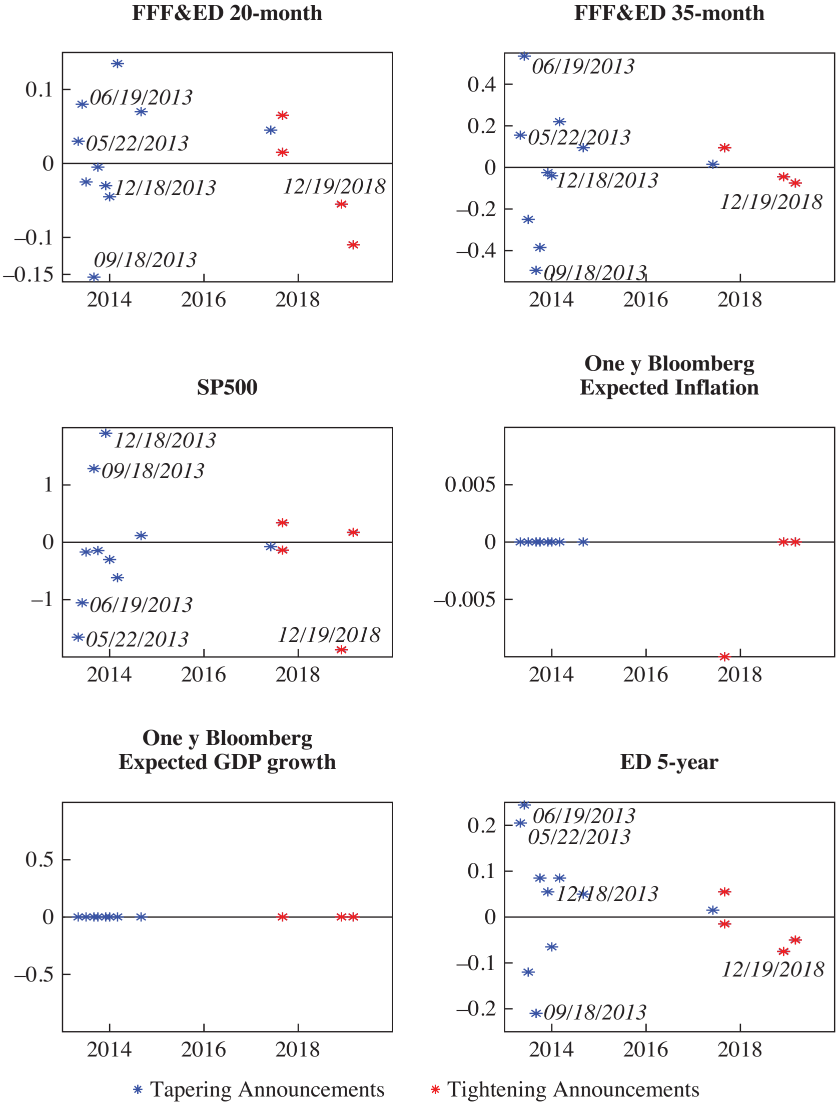
To study the major Fed communications pertaining to its balance sheet normalization plans, I examine communications pertaining to both tapering the pace of its asset purchases, as well as statements pertaining to quantitative tightening (QT), that is, outright reductions in the size of the balance sheet, on grounds that tapering is the first step toward tightening. For brevity, I refer to both types of communication events as QT events.

To identify QT-specific events, we do an exhaustive analysis of published or recorded Fed communications about its balance sheet normalization process dating back to May 2013. The full list of QT events identified is given in the online appendix. We identify fourteen QT-specific events from May 22, 2013, to March 20, 2019, spanning Federal Open Market Committee (FOMC) press releases, Fed chair press conferences, and Fed chair congressional testimonies. We focus on communications that were both specifically about QT and for which QT alone or (in a few cases) QT used in tandem with other types of unconventional monetary policy, such as forward guidance, was the predominant source of news during the short event time window.[2] My focus here will be on the most relevant communications for the stock market. I study changes in market variables from ten minutes before the beginning of the identified QT communication to the close of market trading on the same business day.

We begin by looking at the reaction of high-frequency variables to Fed QT announcements, including minutely observations on the federal funds futures market and on the S&P 500 stock market index, and daily measures of professional forecasts of inflation and GDP growth from Bloomberg. For each of the fourteen QT events in our sample, figure 1 displays the log change in these variables over the high-frequency event windows described above (with the exception of the Bloomberg forecasts which are day before/day after). The five most quantitatively important events for the stock market are labeled. Besides large jumps in the stock market, we see that some QT events are associated with large jumps in longer-horizon federal funds futures rates. What we do not see—in contrast to the broader FOMC event space studied by Bianchi, Ludvigson, and Ma (2022)—is nonnegligible movement in the daily forecasts of inflation and GDP growth in

2. We augment our understanding of the most important pieces of market news surrounding a given communication by conducting a systematic analysis of newspaper reports from Factiva.

Figure 1. High-Frequency Changes in Prices and Expectations



Sources: Tickdata.com; CME Group; and Bloomberg.

Note: For each Fed announcement about balance sheet normalization, the log change in the observed variables in a short time-window around the announcement is shown. In most cases, this corresponds to ten minutes before the announcement to the end of the stock market trading day. For 5/22/2013, 12/18/2013, 9/17/2014, 6/14/2017, 12/19/2018, and 3/20/2019, the twelve-quarter (thirty-six-month) Eurodollar is used in place of missing thirty-five-month federal funds futures data. The labeled dates are the five most quantitatively important Fed announcements based on changes in the S&P 500–lagged GDP ratio, where lagged GDP is the previous month's GDP estimate. The full sample has fourteen balance sheet normalization events spanning May 22, 2013, to March 20, 2019.

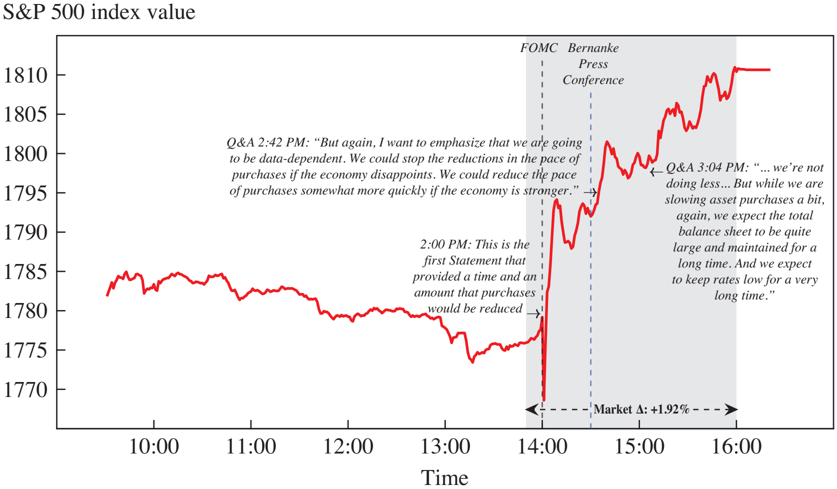
response to the QT events. In particular, the data show that these announcements did nothing to change expectations about the real macroeconomy or inflation, despite large effects on financial market variables.

Figures 2–4 show the intraday movements in the event windows on the three most important QT event dates as measured by the response of the S&P 500 stock market index. These events (from largest to smallest in absolute importance) are: (1) the December 18, 2013, FOMC press release and subsequent Ben Bernanke press conference, which combined to propel the market upward by 1.92 percent from ten minutes before the 2:00 p.m. press release to the end of the trading day (FRB 2013b, 2013c); (2) the December 19, 2018, Jerome Powell press conference in which the market fell 1.9 percent from ten minutes before its beginning at 2:30 p.m. to the end of the trading day (FRB 2018a, 2018b); and (3) the May 22, 2013, Bernanke congressional testimony and subsequent 2:00 p.m. FOMC minutes press release, the combination of which sent the market down 1.7 percent for the day (JEC 2013; FRB 2013a).

In each of the figures 2–4, the shaded area shows the window of time used subsequently to define the QT news event in the structural estimation. In cases where the FOMC press release—in each case at 2:00 p.m.—contained information specifically about balance sheet normalization, the QT event window is measured from ten minutes before the FOMC press release to the close of the stock market. In cases where the post-FOMC press conference contained balance sheet information but the 2:00 p.m. FOMC press did not, the QT event window is measured from ten minutes before the start of the press conference, in each case at 2:30 p.m.

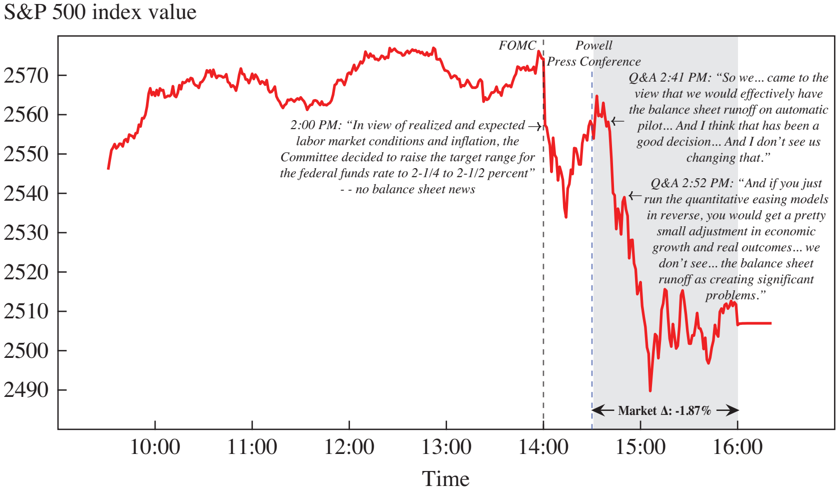
On December 18, 2013, the FOMC made its first statement that provided both when (beginning in January 2014) and by how much the Fed would reduce the pace of asset purchases. An initial drop in the market quickly recovered. News reports indicate that the initial drop was due to the Fed taper statement, but markets recovered when they noticed the taper was tiny—adding to its holdings of mortgage-backed securities (MBS) at a pace of $35 billion per month rather than $40 billion, and to its holdings of Treasury securities at a pace of $40 billion per month rather than $45 billion. The stock market continued to rise during Bernanke's press conference when he emphasized that the Fed would be "data-dependent" (2013c, 5) and flexible with reductions in the pace of purchases and could stop the reductions if the economy disappoints. The market rose further when, in response to questions, Bernanke stated that he expected the balance sheet to be maintained "at a large level for a long time" (17). This is the first of several statements in our sample indicating that the market

Figure 2. Intraday Movements in the S&P 500: December 18, 2013



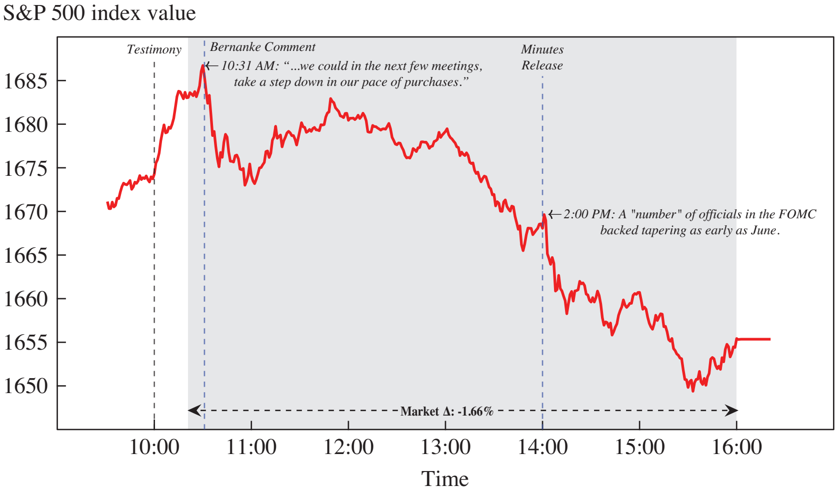
Sources: Tickdata.com; FRB (2013b; 2013c, 7 n1, 17).
Note: The gray shaded area represents the event window used for the high-frequency structural event study.

Figure 3. Intraday Movements in the S&P 500: December 19, 2018



Sources: Tickdata.com; FRB (2018a; 2018b, 6, 11).
Note: The gray shaded area represents the event window used for the high-frequency structural event study.

Figure 4. Intraday Movements in the S&P 500: May 22, 2013



Sources: Tickdata.com; FRB (2013a); FOMC (2013).

Note: The gray shaded area represents the event window used for the high-frequency structural event study.

reacted positively to commentary, suggesting a flexible, "data-dependent" approach to its balance sheet normalization plans, and conversely, as we'll see next, reacted negatively to commentary suggestive of inflexibility.

The most prominent QT event showcasing the converse case was Powell's December 19, 2018, press conference at 2:30 p.m., following an FOMC press release at 2:00 p.m. The FOMC press release contained no news about the balance sheet, hence it is excluded from the QT event window. The market's direction turned downward dramatically during the press conference at a time stamp that immediately followed Powell's responses in the Q&A in which he stated that the FOMC "came to the view that we would effectively have the balance sheet runoff on *automatic pilot*. . . . And I think that has been a good decision. . . . And I don't see us changing that" (FRB 2018b, 6; emphasis added). Press reports suggest that the perceived inflexibility of the "automatic pilot" language was the antithesis of the "data-dependent" commentary of the December 18, 2013, event, with opposite consequences for the stock market. The market declined further when Powell suggested that the Fed did not see the shrinking of its balance sheet as a source of economic instability: "And if you just run the quantitative easing models in reverse, you would get a pretty small adjustment in economic growth and real outcomes. . . . We don't see . . . the balance sheet runoff as creating significant problems" (11).

The May 22, 2013, congressional testimony Bernanke gave resulted in the so-called taper tantrum. In the case of the stock market, the declines began immediately after a 10:31 a.m. Bernanke comment that "we could in the next few meetings, take a step down in our pace of purchases" (JEC 2013, 11). Although the stock market partially recovered later in the day, losses for stocks accelerated once more after the 2 p.m. release of FOMC minutes stating that "a number" of officials in the FOMC supported tapering as early as their next meeting in June (FOMC 2013, 7). News analysis suggests that the stock market focused especially on the information about tapering in the minutes due to Bernanke's comments earlier that morning.

II. Why Did the Market React? Mixed-Frequency Structural Approach

I now use a structural model to make inferences on why the stock market reacted to these QT announcements. I apply the methodology in Bianchi, Ludvigson, and Ma (2022). This section provides a brief description of the model and estimation approach.

Bianchi, Ludvigson, and Ma (2022) integrate a high-frequency monetary event study into a mixed-frequency macro-finance model and structural estimation. We examine Fed communications alongside both high- and lower-frequency data through the lens of a structural equilibrium asset pricing model with New Keynesian–style macroeconomic dynamics. This approach allows us to estimate jumps in investor beliefs about the latent state of the economy, the perceived sources of economic risk, and the future conduct of monetary policy at high frequency surrounding Fed news events. I focus on this aspect of the empirical approach in Bianchi, Ludvigson, and Ma (2022), applied in this instance to Fed announcements about its balance sheet normalization plans.[3]

The main elements of this model are as follows: (1) It is a two-agent model with New Keynesian macro dynamics with heterogeneous beliefs. Households or workers invest in short-term bonds but have no stock

3. The mixed-frequency structural estimation further permits us to quantify the causal effects of shifts in monetary policy that may occur outside of tight windows surrounding Fed communications. The interested reader can find this analysis in Bianchi, Ludvigson, and Ma (2022).

market wealth; their expectations are formed using backward-looking rules. Investors are forward-looking, and they can react quickly to news. Their expectations are consistent with an understanding of driving forces in the model but they must form beliefs about the future conduct of monetary policy. They earn all income from investments in risk-free nominal bonds and the stock market. (2) The conduct of monetary policy is subject to infrequent nonrecurrent regime shifts, or structural breaks, that take the form of shifts in the parameters of a nominal interest rate rule. Investors understand that breaks occur and form an expectation of what policy rule will come next, once the current regime ends. (3) There are six primitive shocks: a monetary policy shock in the interest rate rule; an aggregate demand shock in the real activity/IS equation; a markup shock in the Phillips curve; a shock to trend growth; an earnings share shock that redistributes the rewards of production between workers and investors without affecting the size of the rewards; and a liquidity premium that represents a preference for risk-free nominal debt over equity. This captures exogenous movements in the equity premium that could be attributable to fluctuations in the liquidity and safety attributes of risk-free nominal debt (Krishnamurthy and Vissing-Jorgensen 2012), changes in risk aversion, flights to quality, and jumps in sentiment. (4) We estimate jumps in investor beliefs around Fed news events about the current economic state ("nowcasts"), about the perceived sources of economic risk, and about future regime change in the monetary policy rule. (5) Numerous forward-looking series at mixed frequency are used to map theoretical implications for beliefs, markets, and the economy into data. (6) The full structural model is solved and estimated using Bayesian methods.

To understand the impetus for modeling two types of agents (households versus investors), note that household survey data indicate that households display substantial inertia in the expectations (Malmendier and Nagel 2016).[4] On the other hand, it is evident that financial markets react swiftly to central bank communications and actions. This suggests that the expectations of financial market participants are subject to little inertia. The framework reconciles these seemingly contradictory observations by considering two types of agents with different beliefs.

It is worth discussing the channels through which quantitative interventions could influence the model economy. To do so, it is helpful to

4. We follow Malmendier and Nagel (2016) in modeling household inflation expectations as evolving from a constant gain learning algorithm, and we discipline our estimates of the parameters of this process by filtering household expectations data from the University of Michigan Survey of Consumers.

present two equations. Equation (1) is the central bank's interest rate policy rule, which takes the general form:

$$(1) \qquad i_t - \left(r_{ss} + \pi^T_{\zeta_t}\right) = \left(1 - \rho_{i,\zeta_t}\right)\left[\psi_{\pi,\zeta_t}\left(\pi_t - \pi^T_{\zeta_t}\right) + \psi_{\Delta y,\zeta_t}\left(y_t - y_{t-1}\right)\right]$$
$$+ \rho_{i,\zeta_t}\left[i_t - \left(r_{ss} + \pi^T_{\zeta_t}\right)\right] + \sigma_i \varepsilon_{i,t}, \varepsilon_i \sim N\left(0,1\right),$$

where i_t is the short-term nominal interest rate, r_{ss} is the steady-state real interest rate, π_t is current inflation, y_t is aggregate output, $\varepsilon_{i,t}$ is a monetary policy shock, and $\pi^T_{\zeta_t}$, ψ_{π,ζ_t}, $\psi_{\Delta y,\zeta_t}$, ρ_{i,ζ_t} are time-varying parameters of the policy rule where ζ_t denotes a discrete-valued random variable that indexes the estimated policy regimes in our sample. Lags of the variable on the left-hand side appear in the rule to capture the observed smoothness in adjustments to the central bank's target interest rate.

Equation (2) is the log equity premium as perceived by the investor:

$$(2) \qquad \underbrace{E^b_t\left[r^D_{t+1}\right] - \left(i_t - E^b_t\left[\pi_{t+1}\right]\right)}_{subj.\ equity\ premium}$$
$$= \underbrace{\left[-.5V^b_t\left[r^D_{t+1}\right] - COV^b_t\left[m_{t+1}, r^D_{t+1}\right] - .5V^b_t\left[\pi_{t+1}\right] - COV^b_t\left[m_{t+1}, \pi_{t+1}\right]\right]}_{subj.\ risk\ premium} + \underbrace{lp_t}_{liquidity\ premium}$$

where E^b_t [.] V^b_t [.], and COV^b_t [.] are the conditional mean, variance, and covariance under the subjective beliefs of the agent. The equity premium has two components. The component labeled "subj. risk premium" is the part attributable to the agent's subjective perception of risk. As explained in Bianchi, Ludvigson, and Ma (2022), this component is driven entirely by realized regime changes in the conduct of monetary policy or investors' subjective beliefs about the probability of a near-term regime change in the policy rule. The liquidity premium is a catchall for all sources of time variation in the equity premium other than those attributable to shifts in subjective beliefs about the monetary policy rule.[5]

With these equations in mind, we can discuss the channels through which quantitative interventions could influence the model economy. First, monetary policy is summarized by the interest rate rule, equation (1), thus the framework doesn't explicitly model quantitative interventions in the form

5. In our structural estimation we use the twenty-year BAA-Treasury spread as a noisy signal of this component.

of explicit Fed purchases of long-term Treasuries, agency debt, or agency MBS. However, quantitative interventions and other forms of unconventional monetary policy, such as forward guidance, show up in the policy rule implicitly through their influence on the time-varying parameter $\pi^T_{\xi_t}$. Although this parameter plays the role of an inflation target in the interest rate rule, unlike traditional New Keynesian models, $\pi^T_{\xi_t}$ does not necessarily coincide with the stated long-term inflation objective of the central bank. This happens because the model here differs in two ways from the traditional New Keynesian models: macro (household) expectations are estimated to be strongly backward-looking—implying that long-term inflation expectations of households can persistently deviate by large magnitudes from $\pi^T_{\xi_t}$ even though they eventually converge toward $\pi^T_{\xi_t}$—and because the policy rule parameters are not constant but instead vary over time. In this setting, $\pi^T_{\xi_t}$ is more appropriately thought of as an implicit time t target rather than an explicit objective. Forward guidance and quantitative easing (QE), two tools that were employed at the zero lower bound, are channels that manifest indirectly in the policy rule as a higher value for the implicit inflation target $\pi^T_{\xi_t}$, since these policies are designed in part to generate higher expected inflation and lower real rates (thereby stimulating aggregate demand) even as nominal interest rates remain unchanged at the zero lower bound. We could thus refer to this as the inflation expectations channel of unconventional monetary policy transmission. Movements in the real interest rate are the primary channel of monetary transmission to the aggregate macroeconomy in the model of Bianchi, Ludvigson, and Ma (2022) and in New Keynesian models in general. However, in our model such unconventional monetary interventions must be effective at actually changing inflation expectations in order for this channel to be operative. Because we estimate that household inflation expectations respond only very slowly over time to new information about inflation and changes in the implicit target $\pi^T_{\xi_t}$, we estimate that this inflation expectations channel is quite muted, which we stress is a result rather than an assumption.

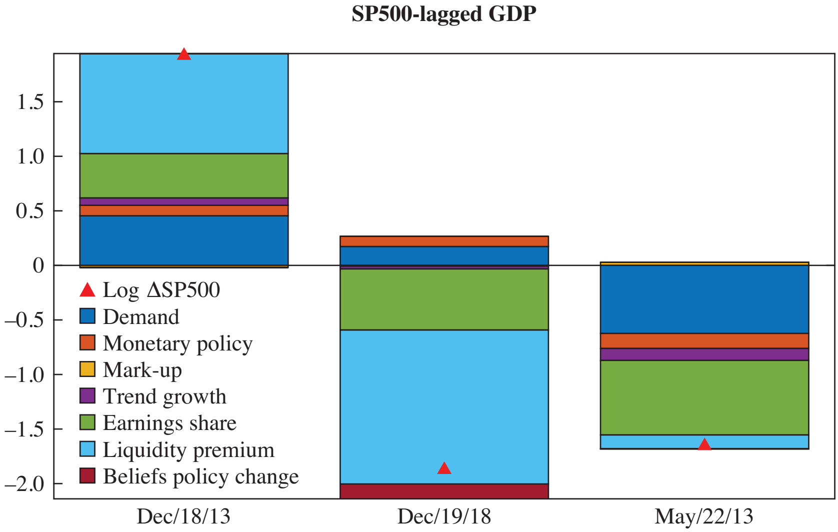
Yet even if quantitative interventions have a limited effect on the macroeconomy through the inflation expectations channel, these interventions—and the Fed's announcements about them—could still have quantitatively important effects on financial markets through three other distinct channels in the model: (1) by distorting return premia in financial markets; (2) by altering investor beliefs about broader economic activity, such as output growth or inflation (the "Fed information effect"); or (3) by affecting investor nowcasts of the share of output accruing to equity holders (as opposed to workers). The first channel can be triggered if Fed news about balance sheet normalization either changes the liquidity premium or causes jumps in beliefs about a near-term regime shift in the policy rule. Beliefs about regime change in the policy rule play a crucial role in shaping perceptions of equity market risk. For example, a jump upward in the perceived probability of a near-term shift to a policy rule with greater activism in stabilizing the real economy (manifested as larger values for the activism coefficient $\psi_{\Delta y,\zeta_t}$) lowers expected volatility, driving the subjective risk premium down and the stock price up. Each of these channels can have large effects on the stock market in the model economy, but unless they are accompanied by changes in the real interest rate—through the inflation expectations channel—quantitative interventions will have no effect on the broader macroeconomy in the model.

For the subperiod that is relevant for balance sheet normalization, our estimates imply that markets were expecting the next policy rule to be both more hawkish (lower $\pi^T_{\zeta^P_t}$) and more active, with higher values for both ψ_{π,ζ_t} and $\psi_{\Delta y,\zeta_t}$. These two forces have offsetting effects on stock market valuations. The expectation of a more hawkish Fed works to lower the stock market's value by raising the perceived probability of persistently higher real rates. By contrast, the expectation of a more active Fed would work to raise the stock market's value by lowering the perceived quantity of risk in the market. Our estimates imply that Fed announcements in this period have a larger effect on the perceived quantity of risk than they do on the path of future short rates, so that Fed communications that trigger a lower perceived probability of transitioning to the next policy rule decrease the stock market's value on net.

III. High-Frequency Structural Analysis: What Did the Market Learn?

What did markets learn from these QT events? I use the methodology of Bianchi, Ludvigson, and Ma (2022), which combines a filtering algorithm with a structural estimation, to decompose movements in forward-looking variables such as the stock market into revisions in beliefs about the primitive shocks affecting the economy and about the possibility of a near-term regime shift in monetary policy. The novelty of this approach allows us to investigate a variety of possible explanations for why markets respond strongly and swiftly to central bank actions and announcements, not merely by delineating which expectations are revised but also

Figure 5. Sources of Movements in the S&P 500: Top Three QT Events



Source: Author's calculations.

Note: The figure reports the decomposition of movements in the S&P 500–lagged GDP ratio attributable to revisions in the perceived shocks hitting the economy and in the belief regimes for the five most quantitatively important Fed announcements (as measured by the absolute magnitude of jumps in the S&P 500–lagged GDP ratio) about balance sheet normalization.

by providing granular detail on why they are revised, with a decomposition of market responses into the primitive economic sources of risk responsible for observed forecast revisions.

Figure 5 shows the decomposition of jumps in the S&P 500–lagged GDP ratio into components driven by different elements of the perceived vector of Gaussian shocks and by investor beliefs about the probability of a regime shift in the monetary policy rule. These are estimates of how investors' perceived shocks were revised due to the Fed news. For example, if the confluence of data suggests that stock market investors learn from an announcement that there has been a restrictive monetary policy shock, this shows up in our structural estimation as a negative contribution to the stock market. If, at the same time, investors have revised their nowcasts for aggregate demand up—that is, they perceive a higher demand shock than previously as a result of the announcement—this shows up in our structural estimation as making an offsetting positive contribution. Note that jumps in the S&P 500–lagged GDP ratio at QT announcements are entirely attributable to jumps in the stock market, since GDP is lagged one month.

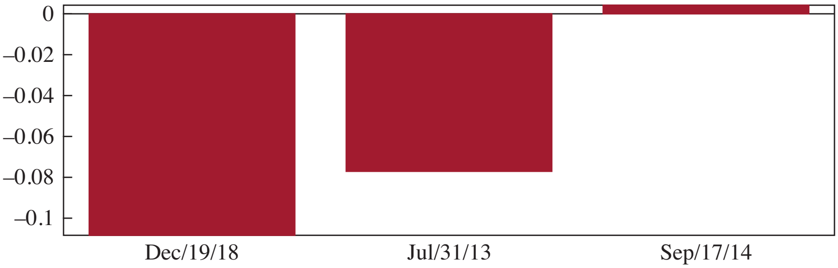
The triangles in the figure mark both the actual change in the S&P 500 during the event window, as well as the model-implied change in the stock market in the window. These two coincide exactly, since our state-space estimation disallows observation errors in the observation equation for the S&P 500–lagged GDP ratio.

Figure 5 shows that the most quantitatively important QT event in our sample—December 18, 2013, when the market rose 1.92 percent in the two hours surrounding the news—was largely driven by a lower nowcast for the liquidity premium component of the subjective equity premium and higher investor nowcasts for the earnings share of output and for aggregate demand, with small supporting contributions from the perception of a more accommodative monetary policy shock and higher trend growth. The second most important QT event for the stock market was on December 19, 2018, when the market fell 1.9 percent in the ninety minutes surrounding Powell's remarks. In this case, the lion's share of the decline was attributable to the subjective equity premium rising, mostly due to the liquidity premium rising, but also due to a jump downward in the perceived probability of a near-term regime change in the conduct of monetary policy. In addition, the nowcast for the earnings share fell, contributing to the decline.

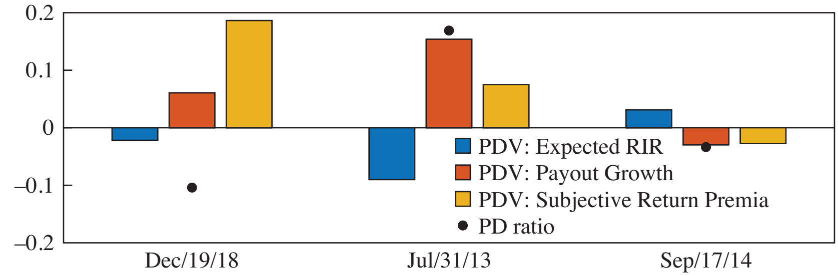
If we sort events according to their importance for revisions in investor beliefs about the probability of regime change in the policy rule, we find that the QT event of Powell's press conference on December 19, 2018, is by far the most important. Figure 6, panel A, shows the change in the perceived probability of a regime change for each of these five events, while panel B shows the decomposition of the jump in the model price–payout ratio, pd_t, into its various contributing forces (subjective return premia, expected real interest rates, and expected payout growth). Panel A shows that the December 19, 2018, QT event is associated with a large downward revision in the perceived probability of a regime change in the policy rule. Panel B shows that this same event is associated with a jump downward in pd_t (the dot), driven almost entirely by a large jump upward in subjective expected return premia. Subjective perceptions of risk rise, in part, because of the sharp decline in the perceived probability of transitioning to the policy rule expected to come next, where the central bank would be more actively engaged in stabilizing the real economy. The decline in the perceived probability of transitioning to this next rule raises expected volatility and the subjective equity premium. This is the structural interpretation of Powell's "automatic pilot" comment about runoff, seen through the lens of the model.

Figure 6. Jumps in Risk Perceptions, Short Rates, and Earnings Expectations

Panel A: Change in perceived probability of a policy regime change



Panel B: Present discounted values (*pdv*): reinvestment ratio (*rir*), earning growth, and excess returns



Source: Author's calculations.

Note: Panel A shows the pre-announcement and post-announcement changes in the perceived probability that financial markets assign to a switch in the monetary policy rule occurring within one year, for the five most quantitatively important QT announcements based on changes in investor beliefs about the future conduct of monetary policy. Panel B shows a decomposition of the model's fluctuations in the log price–payout ratio $pd = pdv_t(\Delta d) - pdv_t(r^{ex}) - pdv_t(rir)$ in the same event windows around these announcements that are driven by subjective equity risk premium variation, as measured by $pdv_t(r^{ex})$; subjective expected future real interest rate fluctuations, as measured by $pdv_t(rir)$; and subjective expected earnings growth, as measured by $pdv_t(\Delta d)$. PD ratio is $pdv_t(\Delta d) - pdv_t(r^{ex}) - pdv_t(rir)$.

IV. Taking Stock

What do we take away from these results? Two points stand out. First, whether it's tapering or tightening, the stock market dislikes perceived inflexibility in statements about balance sheet normalization. Second, financial markets, including the stock market, are clearly attuned to news about the balance sheet. And the stock market is reactive to such news. The subjective equity return premium is a big driver of jumps in the market around QT news events, with the jumps in nowcasts for the earnings share playing an important secondary role. Whether these stock market moves in response to QT news have any implication for the broader economy is an open question. We find little evidence that high-frequency measures of forecasts of inflation or real GDP growth respond at all to QT news events, despite the large stock market reactions. Extensive literature on asset pricing suggests that much of the variation in stock market return premia has a negligible correlation with broader economic activity. Finally, the movements in the earnings share that we measure—an important source of variation in the stock market—merely redistribute the rewards of production without affecting the size of those rewards. Thus, by construction, perceived changes in this share in response to Fed news have nothing to do with expectations for the broader economy.

ACKNOWLEDGMENTS This discussion is based on joint work with Francesco Bianchi and Sai Ma. We are grateful to Courtney Wiegand, Nicholas Zarra, and Steven Zheng for excellent research assistance.

References

Bianchi, Francesco, Sydney C. Ludvigson, and Sai Ma. 2022. "Monetary-Based Asset Pricing: A Mixed-Frequency Structural Approach." Working Paper 30072. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30072.

Federal Open Market Committee (FOMC). 2013. "Minutes of the Federal Open Market Committee April 30–May 1, 2013." https://www.govinfo.gov/content/pkg/CHRG-113shrg81472/pdf/CHRG-113shrg81472.pdf.

Federal Reserve Board (FRB). 2013a. "Minutes of the Federal Open Market Committee." Press release, May 22. https://www.federalreserve.gov/newsevents/pressreleases/monetary20130522a.htm.

Federal Reserve Board (FRB). 2013b. "Federal Reserve Issues FOMC Statement." Press release, December 18. https://www.federalreserve.gov/newsevents/pressreleases/monetary20131218a.htm.

Federal Reserve Board (FRB). 2013c. "Transcript of Chairman Bernanke's Press Conference, December 18, 2013." https://www.federalreserve.gov/mediacenter/files/fomcpresconf20131218.pdf.

Federal Reserve Board (FRB). 2018a. "Federal Reserve Issues FOMC Statement." Press release, December 19. https://www.federalreserve.gov/newsevents/pressreleases/monetary20181219a.htm.

Federal Reserve Board (FRB). 2018b. "Transcript of Chairman Powell's Press Conference, December 19, 2018." https://www.federalreserve.gov/mediacenter/files/fomcpresconf20181219.pdf.

Joint Economic Committee (JEC). 2013. "Hearing before the Joint Economic Committee Congress of the United States," May 22. Washington: US Government Printing Office. https://www.govinfo.gov/content/pkg/CHRG-113shrg81472/pdf/CHRG-113shrg81472.pdf.

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

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

JONATHAN H. WRIGHT
Johns Hopkins University

The Extent and Consequences of Federal Reserve Balance Sheet Shrinkage

ABSTRACT This paper discusses the process of balance sheet shrinkage that the Federal Reserve is currently undertaking. I argue that the overall balance sheet is unlikely to shrink by much and that it will remain a much larger share of nominal GDP than it was before the COVID-19 pandemic. I examine the effects of balance sheet shrinkage on asset prices, taking the perspective that these effects are mostly likely to be narrow, that is, specific to the price of the asset that the market has to absorb rather than spilling over to fixed income prices more generally. I argue that the effects of reducing the Fed's holdings of Treasuries can be thought of as equivalent to the Treasury increasing the amount and maturity of its issuance. I estimate that this will have very small effects on term premia and bond yields. The reduction of the Fed's holdings of mortgage-backed securities might have larger effects on the yields of these securities, especially if the Fed starts selling these securities. Any substantive macroeconomic effect of balance sheet runoff is likely to operate through mortgage rates and the housing market.

In May 2022, the Federal Open Market Committee (FOMC) announced plans for shrinking the size of the balance sheet, a plan often referred to as quantitative tightening (QT). This did not call for any outright asset sales but rather limiting reinvestment of maturing assets. The program was phased in over three months but has now reached its full extent, according to which Treasuries are reinvested only to the extent that they exceed a

Conflict of Interest Disclosure: Jonathan H. Wright is a visiting scholar at the Federal Reserve Bank of Philadelphia. Other than the aforementioned, 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, Fall 2022: 259–275 © 2023 The Brookings Institution.

$60 billion per month cap; for principal repayments of mortgage-backed securities (MBS), the corresponding cap is $35 billion. Most of the MBS in the system open market account (SOMA) portfolio are at low coupon rates and have low prepayment speeds given that homeowners will not choose to refinance. As such, the actual pace of shrinkage of MBS is likely to be much less than $35 billion. Ennis and Kirk (2022) projected a pace slightly above $20 billion per month over the next two years.[1] Still, the total pace of Fed balance sheet shrinkage is about twice as fast as it undertook in 2017–2019. The Federal Reserve is also applying the $60 billion per month of Treasury redemptions first to coupon securities and then to bills. This will gradually lower the weighted average maturity of SOMA Treasury holdings.[2]

This paper discusses the likely effects of the program of balance sheet shrinkage, starting with the likely extent of QT and followed by its asset market and macroeconomic impacts. Throughout, I am thinking of the impacts of QT as the difference in outcomes (such as Treasury yields) with the balance sheet actually chosen by the Fed relative to what would have occurred if the Fed were instead to keep the balance sheet constant as a share of nominal GDP at its peak level of 37 percentage points.

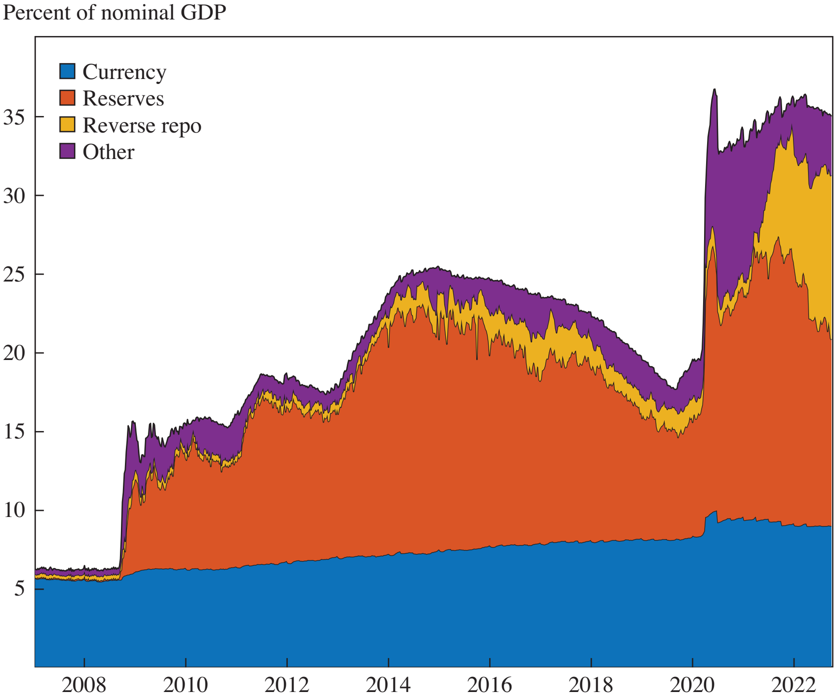
I. The Extent of QT

The implementation of monetary policy changed dramatically in the wake of the global financial crisis and the authorization that Congress gave the Fed to pay interest on reserves. In contrast to the old system of monetary policy implementation with scarce reserves, which is unfortunately still often taught in intro macroeconomics classes (Ihrig and Wolla 2020), the Fed now sets the interest rate on reserves which puts a floor on banks' reserve demand, and then the Fed supplies an amount of reserves that ensures that equilibrium is always on the flat part of the reserve demand curve. Since only depository institutions are eligible to receive interest on reserves, it turns out that this can lead to segmentation whereby short-term interest rates are generally well below the level of interest on reserves. To counter this, the Fed has introduced a system of reverse repos which allow the Fed to effectively pay interest to other entities, such as money market

1. Their assumption, writing in spring 2022, for the terminal level of thirty-year mortgage rates was 5 percent, and by fall, rates had already soared well beyond that, but as they note, the MBS in question have such low coupons that refinancing is unattractive in any case.

2. At present, the weighted average maturity of the SOMA Treasury holdings is 8.3 years, whereas that of marketable Treasuries outstanding is 6.2 years.

Figure 1. Federal Reserve Liabilities



Sources: US Bureau of Economic Analysis; Board of Governors of the Federal Reserve System.

Note: The liabilities are from the Federal Reserve, "Factors Affecting Reserve Balances—H.4.1," https://www.federalreserve.gov/releases/h41/ release; nominal GDP is from FRED, https://fred.stlouisfed.org/series/GDP.

mutual funds and government-sponsored enterprises. Thus, in effect, the Fed supplies ample reserves and puts two floors on interest rates, one via interest on reserves and the other via interest on reverse repos.[3]

Figure 1 shows the evolution over time of the liabilities of the Federal Reserve System, all scaled by nominal GDP. The size of the Fed's balance sheet soared after the financial crisis and the subsequent shrinkage was quite limited.[4] During the COVID-19 pandemic, the balance sheet expanded again to a peak of 37 percent of GDP and has now begun to shrink.

3. See Dawsey, English, and Sack (forthcoming) for a clear exposition of the Fed's new implementation system.

4. I am defining the size of the Fed balance sheet as the line in the H.4.1 release labeled "Total factors supplying reserve funds," which is a bit bigger than "Securities held outright" because the Fed has assets other than securities.

As the Fed shrinks the balance sheet, it has made it clear that it intends to keep this new system of monetary policy implementation but with as small a balance sheet as possible. In May 2022, the Federal Reserve Bank of New York laid out projections for the process of balance sheet shrinkage (Federal Reserve Bank of New York 2022). They projected that the balance sheet would shrink to 22 percent of nominal GDP, consisting of bank reserve balances of 8 percent of GDP, negligible reverse repos, and 14 percent of nominal GDP in currency and other liabilities (so-called autonomous factors). At the current pace of balance sheet shrinkage, this would continue until 2025 with a Fed balance sheet of about $5.9 trillion before growth would resume.[5]

I am very skeptical that the Fed will ultimately shrink the balance sheet by anything like that much. Bank reserve balances of 8 percent of GDP would get very close to the point where, as in September 2019, banks ended up being on the steep part of their reserve demand curve, causing sharp spikes in the federal funds rate. Copeland, Duffie, and Yang (2021) find that both before and after the spike in September 2019, there were strains in intraday payments that can be tied directly to a shortage of reserves—the sharp rise in rates was an extreme manifestation of a broader shortage of reserves.[6] Furthermore, there is reason to believe that the kink in bank reserve demand is now at a higher level relative to nominal GDP than it was in September 2019. For one thing, since the end of 2019, whereas nominal GDP has risen 16 percent, the total assets of commercial banks in the United States have climbed 28 percent, and that seems a more natural way of scaling reserve demand (Afonso and others 2022). Afonso and others (2022) estimate a nonlinear reserve demand curve using a time-varying vector autoregression and find that the curve has shifted upward and to the right over time. Having bank reserves at around 10 percent of nominal GDP seems a more likely steady state and is close to, but a bit below, the estimates of Afonso and others (2022).[7]

In September 2022, the Fed was offering an overnight reverse repo facility at a fixed offering rate that is 10 basis points below the rate of interest on

5. The projections assume a period of tapering at the end of QT, but I abstract from this for simplicity.

6. The Fed has put in place a standing repo facility as a backstop to prevent spikes in rates, but it is not being used much and may well end up subject to the same problem of stigma that has arisen with the discount rate.

7. Afonso and others (2022) scale reserve demand by bank assets and estimate that the reserve demand curve now becomes steep with reserves below 13–14 percent of bank assets, which corresponds to about 11–12 percent of nominal GDP.

reserves, and the usage of this facility was around 10 percent of nominal GDP, as shown in figure 1. The Fed can, of course, make this facility less attractive by widening the interest spread relative to interest on reserves or by restricting access. But that then weakens its ability to control short-term interest rates and runs the risk that the interest paid to commercial banks on reserves will be substantially above the overall level of short-term interest rates. Moreover, once the Fed shrinks the balance sheet to the point that it does not have the capacity to offer a large overnight reverse repo facility, there is no ready way of going back to restart it again.[8] For these reasons, I think that the Fed will keep room on its balance sheet for a reverse repo facility of at least 6 percent of nominal GDP, which would still involve shrinking it substantially from its current size.

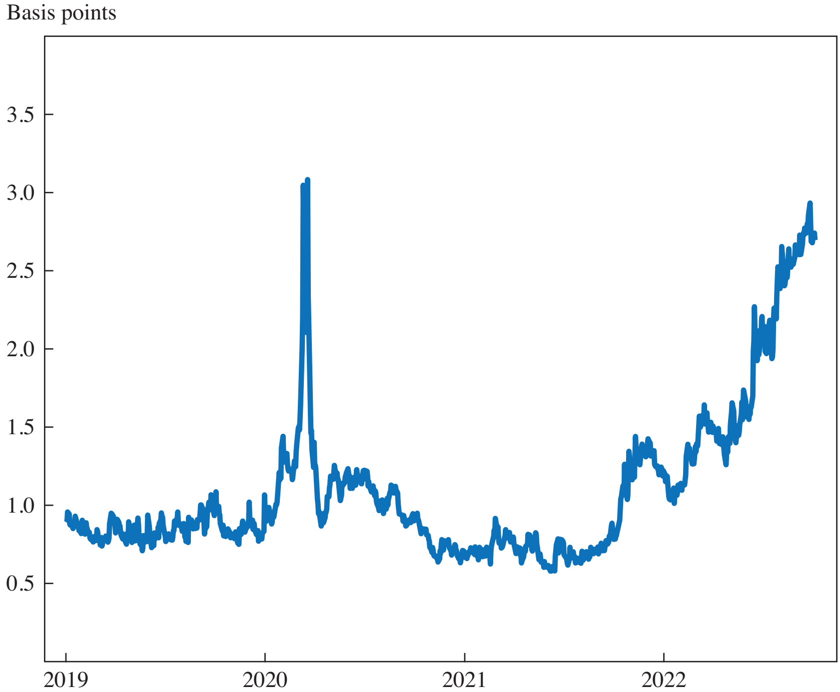
Lopez-Salido and Vissing-Jorgensen (2022) estimate the relationship between the spread of the effective federal funds rate less interest on reserves and the size of reserves plus overnight reverse repos. They conclude that the sum of reserves and overnight reverse repos can be reduced to 15.3 percent of nominal GDP, while avoiding daily spikes. This is almost identical to the 16 percent that I am assuming. With autonomous factors at 14 percent of nominal GDP, as in the SOMA projections, all this implies a steady-state level of the Fed balance sheet of about 30 percent of nominal GDP. At the current pace of balance sheet shrinkage and with the Survey of Professional Forecasters projections for nominal GDP, this will be attained in the middle of 2023 at a level a bit below $8 trillion.[9] And while I can certainly see circumstances in which the balance sheet shrinkage would proceed further, there are also circumstances in which it could end earlier still. Measures of liquidity in the Treasury market are rather poor at the moment.[10] Figure 2 shows the average absolute fitting

8. In the reverse repo facility, the Fed transfers a security to a counterparty, that counterparty deposits cash with a bank, and that bank's reserves with the Fed are debited. The transaction is unwound the next day. Thus the reverse repo extinguishes reserves. If reverse repo facilities are largely ended and the balance sheet is shrunk so that it is equal to the sum of autonomous factors and the minimum level of reserves demanded by banks on the flat part of the demand curve, then subsequently restarting reverse repos in size in a crisis will create reserve scarcity.

9. Federal Reserve Bank of Philadelphia, "Survey of Professional Forecasters," https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/survey-of-professional-forecasters.

10. One of the autonomous factors that has grown substantially is the Treasury General Account (TGA), which is both high and volatile. Swings in the TGA mechanically cause shifts in reserves. Before the financial crisis, the Treasury would instead hold its cash at private banks. If the Fed could push the TGA back to private banks, then they could keep balance sheet shrinkage going a bit longer.

Figure 2. Average Absolute Treasury Fitting Errors

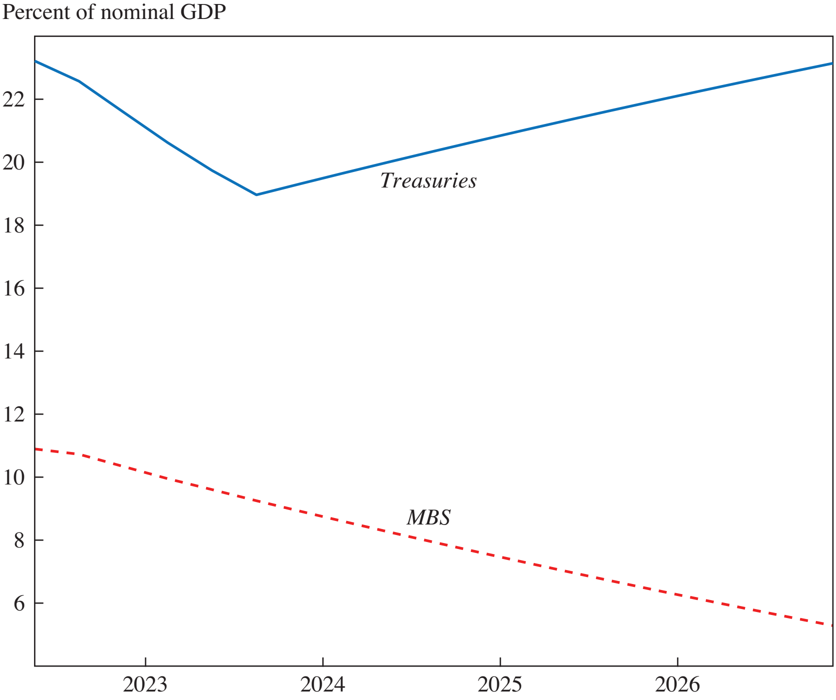


Source: Bloomberg.
Note: This figure shows the mean absolute fitting error of a smoothed yield curve fitted to Treasury securities.

error on Treasury securities—a measure of liquidity proposed by Hu, Pan, and Wang (2013)—and this is at a high level at present. Further deterioration in Treasury market liquidity might lead the Fed to end balance sheet shrinkage early. Demand for the overnight reverse repo facility may remain elevated, and the Fed may be reluctant to do much to restrict its usage. An economic downturn that comes sooner than expected might also lead the process to stop early.

When the Fed does stop shrinking the overall size of its balance sheet, it intends to continue shrinking holdings of MBS with a view to eventually reverting to something close to a Treasuries-only balance sheet, and I assume that the Fed will do this. But that would mean that its holdings of Treasuries would have to expand somewhat faster than nominal GDP to keep the overall balance sheet constant as a share of GDP. All in all, I expect

Figure 3. Projected Assets on Federal Reserve Balance Sheet



Sources: Survey of Professional Forecasters; author's calculations.

Note: The SOMA holdings of Treasuries and MBS for 2022:Q2–2026:Q4 are shown as a percentage of nominal GDP. SOMA holdings for 2022:Q2 and 2022:Q3 are actual; those for subsequent quarters are computed assuming that Treasuries decline by $60 billion per month and MBS decline by $22 billion per month until the balance sheet hits 30 percent of nominal GDP; subsequently MBS continue to decline at that pace while Treasuries hold the balance sheet ratio to GDP constant. Nominal GDP projections are taken from the August 2022 Survey of Professional Forecasters through 2023:Q3, and growth at an annualized rate of 4 percent is assumed thereafter. Note that the Treasuries and MBS sum to a little less than the total size of the Federal Reserve balance sheet because they have other assets as well.

that the decline in the Fed's holdings of Treasuries will be both small (relative to nominal GDP) and transitory. Figure 3 illustrates a prospective path of Treasuries and MBS as shares of nominal GDP computed on the assumption that Treasuries fall by $60 billion per month until the total balance sheet hits 30 percent of GDP and that MBS fall by $22 billion per month for the remainder of the projection period. The decline in Treasuries as a share of GDP gets completely reversed as MBS continue to run off.

The Fed has left open the possibility of outright sales of MBS, and it will have to do this if its goal of holding primarily Treasuries on the Fed balance sheet is to be anything more than an abstract aspiration. But of course if there are outright sales of MBS, then for a given balance sheet size, the Fed has to hold more Treasuries.

There is clearly something of a ratchet effect going on here—the expansion of the balance sheet creates more demand for reserves and other Fed liabilities, meaning that the expansion of the balance sheet that we saw in the global financial crisis never got reversed, and I am arguing that the same is likely to happen again as a result of the response to the pandemic. Nelson (2019) and Smith and Valcarel (2021) show that interest rate responses to declining reserves are steeper than to increasing reserves. Acharya and others (2022) argue that quantitative easing (QE) induces banks to expand their short-term liabilities, and this in turn boosts their demand for reserves—in this model, once the balance sheet expands, it becomes hard to shrink it back. A ratchet effect is something of a downside to QE that probably will and should be taken into account before its use in the future.

The Fed could, of course, decide to make fundamental changes in its monetary policy implementation framework. US regulators, including the Fed but also the Office of the Comptroller of the Currency and the Federal Deposit Insurance Commission, have encouraged banks to hold reserves. They view reserves as the most liquid asset on a bank's balance sheet, preferable even to Treasury securities, which might be hard to sell quickly in a crisis without triggering a destabilizing fire sale (Bush and others 2019). A big balance sheet can support financial stability by expanding the supply of safe short-term assets and diminishing the incentive of the private sector to create assets that purport to be both safe and liquid but are in fact neither (Greenwood, Hanson, and Stein 2016).[11] And yet, the objective of regulators to some extent works at cross-purposes with monetary policy that wants to achieve a smaller balance sheet. Materially shrinking the balance sheet would involve substantial changes and agreement among regulators. Nelson (2019, 2020) argues for going back to something like the scarce reserves regime that existed before the global financial crisis. He highlights costs to a big balance sheet, including the fact that tightening monetary policy with a big balance sheet can lead to the Fed's income being insufficient to cover expenses and a deferred asset being created, an outcome that appears imminent at the time of writing. This clearly has a political and optical cost to the Federal Reserve and is conceivably even a threat

11. McAndrews and Kroeger (2016) also argue that a bigger balance sheet promotes resilience in the payments system.

to its long-run independence (English and Kohn 2022). On the other hand, the current monetary policy implementation framework operates well, in the sense of keeping all short-term rates, including the federal funds rate and repo rates, close to the target with little day-to-day volatility (Duffie and Krishnamurthy 2016; Logan 2019). The decline in the Fed's income is a consequence of long-duration assets being on the balance sheet, which was a choice to stimulate the economy during the pandemic, not a necessary implication of a large balance sheet. The Fed is (wisely, I think) reluctant to make big changes to a monetary policy framework that works, and so I expect them to continue to stick to an ample reserves approach while making mild to moderate efforts to shrink the balance sheet size.

II. The Effects of QT

There is now a large body of literature on understanding the effects of QE. This is made possible in part by the fact that many of the QE announcements came as a surprise and so the event study methodology can identify the impact effects quite precisely. Economists debate whether QE operates through broad channels—affecting the expected path of policy and term premia on all fixed-income assets—or narrow channels, with the price impact limited to the specific security being purchased. While it surely has elements of both, there is now a lot of evidence that narrow or local supply channels are a very large part of the mechanism of QE.[12]

Identifying the effects of QT is more challenging. Because central banks had a long time to prepare for QT, the announcements did not come as big surprises and so the event study methodology is not as powerful, although some authors have looked at the effects of QT announcements (D'Amico and Seida 2020; Smith and Valcarel 2021). And while it is tempting to think of QT as the inverse of QE, there are many reasons why they are quite different. QE happens in part during a time of financial instability; QT occurs during mostly stable financial markets. QE might reinforce forward guidance and affect the expected future path of policy; QT is very explicitly disconnected from the future path of policy. QE might have been understood to signal that more purchases would be undertaken if needed to drive bond prices higher (Haddad, Moriera, and Muir 2022); no such signal

12. See, for example, Krishnamurthy and Vissing-Jorgensen (2011), D'Amico and King (2013), Cahill and others (2013), Joyce and others (2011), Di Maggio, Kermani, and Palmer (2020), and Lucca and Wright (2022).

could possibly be construed from QT. And in the past QE happened at the zero lower bound whereas QT is now occurring away from it, which might make the impact of asset purchases on prices larger (Gagnon and Jeanne 2020). Related to this, D'Amico and Seida (2020) find that balance sheet policy announcements have larger effects when there is more interest rate uncertainty.

To the extent that the effects of QE are thought to operate mainly through narrow channels, it might be reasonable to suppose that the same is true of QT. But otherwise it is hard to draw much inference about the likely impacts of QT from the experience with QE. The Treasury component of QT essentially increases the supply of Treasuries that the arbitrageurs have to absorb—in the framework of Vayanos and Vila (2021)—and also increases the maturity of that supply. As such, we might get some guidance on the likely effects of QT by studying the effects of the supply and composition of Treasury debt during the period before the zero lower bound, as considered by Greenwood and Vayanos (2014). This considers changes in the amount of Treasuries that the market has to absorb without any of the other special features associated with QE.

In an exercise very similar to that undertaken by Greenwood and Vayanos (2014), I regress the ten-year Treasury term premium as estimated by Adrian, Crump, and Moench (2013) onto the maturity-weighted debt-to-GDP ratio of Greenwood and Vayanos (2014).[13] Maturity-weighted debt-to-GDP takes the debt-GDP ratio for each maturity, in decimal form, multiplies it by the maturity in years, and sums them up over all maturities. There is a potential endogeneity problem here in that a larger term premium might motivate debt managers to issue shorter maturity debt. I follow Greenwood and Vayanos (2014) and others by instrumenting the maturity-weighted debt-to-GDP ratio by the unweighted debt-to-GDP ratio, which is purely a function of past fiscal decisions. Table 1 shows the estimates using both ordinary least squares and instrumental variables, which are very similar. Both imply that a one unit increase in maturity-weighted debt-to-GDP increases the ten-year term premium by about 0.34 percentage points. Greenwood and Vayanos (2014) instead regressed the ten-year yield on maturity-weighted debt-to-GDP and controlled for the one-year yield, and they also obtained very similar results. For Treasury securities, we can look up the SOMA holdings and work out both what the maturity-weighted

13. The term premium data are available at Federal Reserve Bank of New York, "Treasury Term Premia," https://www.newyorkfed.org/research/data_indicators/term-premia-tabs. I am grateful to Dimitri Vayanos for providing me with the maturity-weighted debt-to-GDP ratio data.

Table 1. Estimates of the Effects of Maturity-Weighted Debt-to-GDP on Term Premia

	OLS	*IV*
MWGDPt	0.34*	0.32*
	(0.13)	(0.14)
First-stage F		869.00

Source: Author's calculations.

Note: This table reports the results of regressions of the ten-year term premium of Adrian, Crump, and Moench (2013) onto the maturity-weighted debt-to-GDP ratio as calculated by Greenwood and Vayanos (2014) using monthly data from June 1961 to December 2007. The term premium is measured in percentage points. Maturity-weighted debt-to-GDP takes the debt-to-GDP ratio for each maturity, in decimal form, multiplies it by the maturity in years, and sums the results over all maturities. Newey-West standard errors are included in parentheses. Following Lazarus and others (2018), the lag truncation parameter is set to $1.3T^{1/2}$ (rounded to the nearest integer) where T is the sample size.

* $p < .05$, using the nonstandard fixed-b critical values of Kiefer and Vogelsang (2005).

debt-to-GDP ratio is at the start of QT and what it will be at any future date, assuming that QT continues at the current pace and making assumptions about reinvestment decisions. From August 2022 to July 2023 (a plausible end date for QT as discussed above), the maturity-weighted debt-to-GDP ratio held by the SOMA portfolio in Treasuries will decline by 0.25.[14] These Treasury securities have to be absorbed by the market, which would drive the ten-year term premium up but only by about 10 basis points. Of course, the actual scope of QT will become clearer over time: if QT carries on for a longer or shorter period than I assume, then the term premium impact would scale up or down.

II.A. Comparison with Other Estimates

Belton and others (2018), writing in the context of QE and surveying the QE literature, propose a rule of thumb: that adding 1 percent of nominal

14. The calculation is laborious but not complicated. For each month, I take the par value of maturing coupon securities from the Federal Reserve Bank of New York, "System Open Market Account Holdings of Domestic Securities," https://www.newyorkfed.org/markets/soma-holdings. If this amount is less than $60 billion, I assume that the difference is subtracted from the holdings of bills. If it is greater than $60 billion, I assume that the amount of bills is kept fixed and the excess of maturing coupons over the cap are reinvested in coupon securities with an assumed maturity of one hundred months (the approximate weighted average maturity of newly issued nominal notes and bonds). I then compute the exact maturity-weighted value of the portfolio, treating all bills as having a maturity of six months, on August 31, 2022, and July 31, 2023, and scale these by nominal GDP for 2022:Q3 and 2023:Q3 respectively. The nominal GDP numbers are the projections from the August 2022 Survey of Professional Forecasters. This calculation results in the maturity-weighted SOMA Treasury debt-to-GDP ratio declining from 1.73 to 1.48.

GDP to the supply of ten-year equivalent Treasuries raises the term premium by about 6 basis points. As the duration of a ten-year Treasury is about 8.5 years, this rule would say that adding 8.5 percentage points to the maturity-weighted debt-GDP ratio would increase the term premium by 6 basis points. As the increase in the maturity-weighted debt-GDP ratio to be absorbed by the market is 25 percentage points, this corresponds to an 18 basis point increase in the term premium. That's a little larger than my estimate above, but since this rule was calibrated to QE, it may well have larger effects than QT. Crawley and others (2022) use the FRB/US model to estimate the impact of QT on ten-year term premia and get an estimate of around 50 basis points, but this is partly because they are following the New York Fed's SOMA projections, which are for a much bigger extent of QT. If one takes the term premium in Crawley and others (2022) as of mid-2023, when I expect QT will end, it is an increase of about 20 basis points—also in the same ballpark. Wei (2022) estimates the effect of a $2.2 trillion runoff of Treasuries, which is also much bigger than I am assuming, and finds an effect of only 6 basis points.

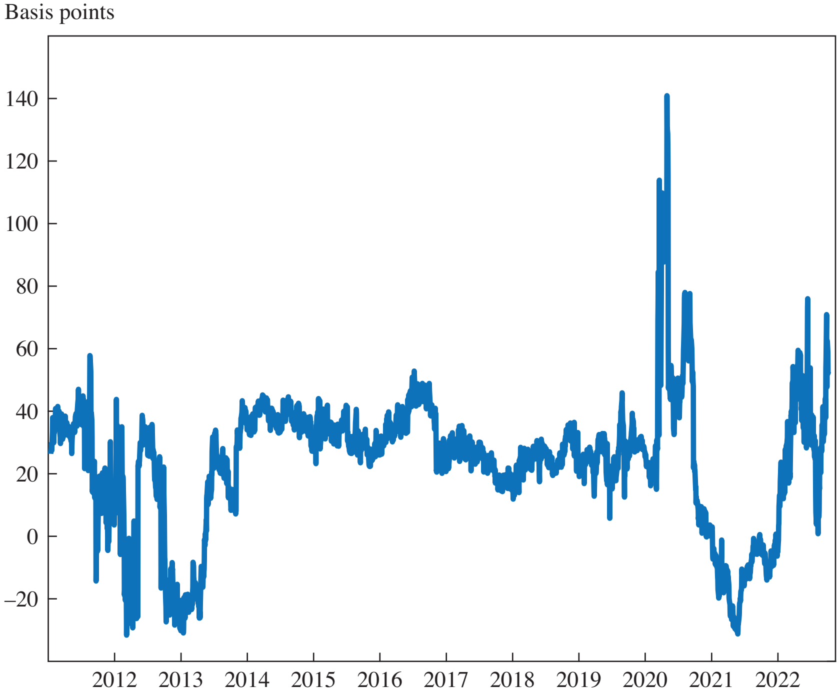
I don't attempt to estimate the macroeconomic impacts of this increase in the term premium, but it is a small and temporary shock to the slope of the yield curve, and generally small and temporary shocks have small macro effects.

II.B. MBS

A different question is the impact of the redemptions of the Fed's MBS portfolio. Unless a new crisis in the housing market develops, the Fed is intent on reducing these holdings permanently and so there may be larger effects here, assuming as I do that the reduction is permanent. Taking the view of asset purchases having predominantly narrow or local effects, any potential impact would be mainly in MBS yields and consequently in mortgage rates and the housing market.

As noted by many authors (for example, Krishnamurthy and Vissing-Jorgensen 2011), the spread of MBS rates over Treasuries, once the MBS yields have been adjusted for their embedded prepayment option, is a natural place to look for the impacts of asset purchases. The current-coupon options-adjusted spread is shown in figure 4. The effects of the QE operations that involved MBS purchases can clearly be seen in this figure: during both the third round of QE and the pandemic, this spread turned negative. In contrast, the limited MBS redemptions in 2017–2019 did not show any effect on this spread. As noted earlier, prepayment rates are

Figure 4. Current Coupon Fannie Mae Option–Adjusted Spread



Source: Bloomberg.
Note: The figure shows the spread between the current coupon Fannie Mae MBS—adjusted for the embedded option—and the corresponding Treasury security.

very slow at the moment because refinancing is uneconomic. Nonetheless, MBS spreads have risen a bit since the QT program reached its full tilt in September. We have very little to go on, but either MBS redemptions are going to go on for a long time or outright sales will begin, and in either case it seems quite plausible that MBS spreads will widen further. At least, this is the place where we should look for material asset price implications of QT. And the effect of MBS balance sheet shrinkage (relative to the counterfactual of holding the Fed's MBS holdings fixed as a share of nominal GDP) is a permanent one. The macroeconomic effect of the MBS spread widening could be important by slowing the housing market, which is already being cooled substantially by the effect of tighter-than-expected conventional monetary policy.

III. Summary and Conclusions

I expect the Fed to resume balance sheet growth with a much higher level of the balance sheet, scaled by nominal GDP, than before the COVID-19 pandemic. The evidence that I have pointed to mostly suggests very small effects on asset prices. If there are to be substantive impacts, I would look to rising MBS spreads adding to the cooling of the housing market.

ACKNOWLEDGMENTS I am grateful to Bill English for helpful comments on an earlier draft. All errors are my sole responsibility.

References

Acharya, Viral V., Rahul S. Chauhan, Raghuram G. Rajan, and Sascha Steffen. 2022. "Liquidity Dependence: Why Shrinking Central Bank Balance Sheets Is an Uphill Task." In *Economic Policy Symposium Proceedings: Reassessing Constraints of the Economy and Policy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Adrian, Tobias, Richard K. Crump, and Emanuel Moench. 2013. "Pricing the Term Structure with Linear Regressions." *Journal of Financial Economics* 110, no. 1: 110–38.

Afonso, Gara, Domenico Giannone, Gabriele La Spada, and John C. Williams. 2022. *Scarce, Abundant, or Ample? A Time-Varying Model of the Reserve Demand Curve*. Staff Report 1019. Federal Reserve Bank of New York.

Belton, Terry, Kristopher Dawsey, David Greenlaw, Huachen Li, Srini Ramaswamy, and Brian Sack. 2018. "Optimizing the Maturity Structure of U.S. Treasury Debt: A Model-Based Framework." Working Paper 46. Washington: Hutchins Center on Fiscal and Monetary Policy at Brookings. https://www.brookings.edu/research/optimizing-the-maturity-structure-of-u-s-treasury-debt/.

Bush, Ryan, Adam Kirk, Antoine Martin, Philip Weed, and Patricia Zobel. 2019. "Stressed Outflows and the Supply of Central Bank Reserves." Blog post, February 20, Liberty Street Economics, Federal Reserve Bank of New York. https://libertystreeteconomics.newyorkfed.org/2019/02/stressed-outflows-and-the-supply-of-central-bank-reserves/.

Cahill, Michael E., Stefania D'Amico, Canlin Li, and John S. Sears. 2013. "Duration Risk versus Local Supply Channel in Treasury Yields: Evidence from the Federal Reserve's Asset Purchase Announcements." Finance and Economics Discussion Series 2013-35. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/econres/feds/duration-risk-versus-local-supply-channel-in-treasury-yields.htm.

Copeland, Adam, Darrell Duffie, and Yilin Yang. 2021. "Reserves Were Not So Ample After All." Working Paper 29090. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29090.

Crawley, Edmund, Etienne Gagnon, James Hebden, and James Trevino. 2022. "Substitutability between Balance Sheet Reductions and Policy Rate Hikes: Some Illustrations and a Discussion." FEDS Notes. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/econres/notes/feds-notes/substitutability-between-balance-sheet-reductions-and-policy-rate-hikes-some-illustrations-20220603.html.

D'Amico, Stefania, and Thomas B. King. 2013. "Flow and Stock Effects of Large-Scale Treasury Purchases: Evidence on the Importance of Local Supply." *Journal of Financial Economics* 108, no. 2: 425–48.

D'Amico, Stefania, and Tim Seida. 2020. "Unexpected Supply Effects of Quantitative Easing and Tightening." Working Paper 2020-17. Federal Reserve Bank of Chicago.

Dawsey, Kristopher, William B. English, and Brian Sack. Forthcoming. "The Federal Reserve Balance Sheet." In *Research Handbook of Financial Markets*, edited by Refet S. Gürkaynak and Jonathan H. Wright. Cheltenham: Edward Elgar.

Di Maggio, Marco, Amir Kermani, and Christopher J. Palmer. 2020. "How Quantitative Easing Works: Evidence on the Refinancing Channel." *Review of Economic Studies* 87, no. 3: 1498–528.

Duffie, Darrell, and Arvind Krishnamurthy. 2016. "Pass-Through Efficiency in the Fed's New Monetary Policy Setting." In *Economic Policy Symposium Proceedings: Designing Resilient Monetary Policy Frameworks for the Future*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

English, William B., and Donald Kohn. 2022. "What If the Federal Reserve Books Losses because of Its Quantitative Easing?" Blog post, June 1, Hutchins Center on Fiscal and Monetary Policy at Brookings. https://www.brookings.edu/blog/up-front/2022/06/01/what-if-the-federal-reserve-books-losses-because-of-its-quantitative-easing.

Ennis, Huberto M., and Kyler Kirk. 2022. "Projecting the Evolution of the Fed's Balance Sheet." *Federal Reserve Bank of Richmond Economic Brief* 22, no. 15.

Federal Reserve Bank of New York. 2022. *Open Market Operations during 2021*. https://www.newyorkfed.org/markets/annual_reports.

Gagnon, Joseph E., and Olivier Jeanne. 2020. "Central Bank Policy Sets the Lower Bound on Bond Yields." Working Paper 20-2. Washington: Peterson Institute for International Economics. https://www.piie.com/publications/working-papers/central-bank-policy-sets-lower-bound-bond-yields.

Greenwood, Robin, Samuel G. Hanson, and Jeremy C. Stein. 2016. "The Federal Reserve's Balance Sheet as a Financial-Stability Tool." In *Economic Policy Symposium Proceedings: Designing Resilient Monetary Policy Frameworks for the Future*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Greenwood, Robin, and Dimitri Vayanos. 2014. "Bond Supply and Excess Bond Returns." *Review of Financial Studies* 27, no. 3: 663–713.

Haddad, Valentin, Alan Moriera, and Tyler Muir. 2022. "Whatever It Takes? The Impact of Conditional Policy Promises." Working Paper. https://conference.nber.org/conf_papers/f170302.pdf.

Hu, Grace Xing, Jun Pan, and Jiang Wang. 2013. "Noise as Information for Illiquidity." *Journal of Finance* 68, no. 6: 2341–82.

Ihrig, Jane, and Scott Wolla. 2020. "Let's Close the Gap: Revising Teaching Materials to Reflect How the Federal Reserve Implements Monetary Policy." Finance and Economics Discussion Series 2020-92. Washington: Board of Governors of the Federal Reserve System.

Joyce, Michael A. S., Ana Lasaosa, Ibrahim Stevens, and Matthew Tong. 2011. "The Financial Market Impact of Quantitative Easing in the United Kingdom." *International Journal of Central Banking* 7, no. 3: 113–61.

Kiefer, Nicholas M., and Timothy J. Vogelsang. 2005. "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests." *Econometric Theory* 21, no. 6: 1130–64.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen. 2011. "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy." *Brookings Papers on Economic Activity*, Fall, 215–65.

Lazarus, Eben, Daniel J. Lewis, James H. Stock, and Mark W. Watson. 2018. "HAR Inference: Recommendations for Practice." *Journal of Business and Economic Statistics* 36, no. 4: 541–59.

Logan, Lorie K. 2019. "Observations on Implementing Monetary Policy in an Ample-Reserves Regime." Speech given to the Money Marketeers of New York University, New York, April 17. https://www.newyorkfed.org/newsevents/speeches/2019/log190417.

Lopez-Salido, David, and Annette Vissing-Jorgensen. 2022. "Reserve Demand and Balance-Sheet Runoff." Working Paper. Washington: Board of Governors of the Federal Reserve System.

Lucca, David, and Jonathan H. Wright. 2022. "The Narrow Channel of Quantitative Easing: Evidence from YCC Down Under." Working Paper 29971. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29971.

McAndrews, James, and Alexander Kroeger. 2016. *The Payment System Benefits of High Reserve Balances*. Staff Report 779. Federal Reserve Bank of New York. https://www.newyorkfed.org/research/staff_reports/sr779.html.

Nelson, Bill. 2019. "The Fed's Balance Sheet Can and Should Get Much Smaller." Remarks given at a Brookings Institution symposium, December 5.

Nelson, Bill. 2020. "The Fed's Precrisis Monetary Policy Framework Is Well-Suited for a Free Society." *Cato Journal* 40, no. 2: 525–32.

Smith, A. Lee, and Victor J. Valcarel. 2021. "The Financial Market Effects of Unwinding the Federal Reserve's Balance Sheet." Working Paper 20-23. Federal Reserve Bank of Kansas City. https://www.kansascityfed.org/research/research-working-papers/financial-market-effects-unwinding-fed-reserve/.

Vayanos, Dimitri, and Jean-Luc Vila. 2021. "A Preferred-Habitat Model of the Term Structure of Interest Rates." *Econometrica* 89, no. 1: 77–112.

Wei, Bin. 2022. "Quantifying 'Quantitative Tightening' (QT): How Many Rate Hikes Is QT Equivalent To?" Working Paper 2022-08. Federal Reserve Bank of Atlanta. https://www.atlantafed.org/research/publications/wp/2022/07/15/08--rate-hikes-and-quantitative-tightening.

General Discussion

Arvind Krishnamurthy commented on the signaling effects described by Sydney Ludvigson and noted that, from previous research, we know that quantitative easing (QE) has important implications for the signal of the path of the policy rate. He interpreted Ludvigson's results as contributing additionally to this by showing that quantitative tightening (QT) announcements change the way in which the market interprets the aggressiveness of the Federal Reserve's policy rate reaction function. On Jonathan Wright's analysis, Krishnamurthy commented that while the Fed ramps up QE rapidly, it exits very slowly—the ratchet effect that Viral Acharya has written about. Tying QE to policy rates, he explained that as long as policy rates are kept low, the balance sheet will be kept high and that around the world this is what central banks have roughly done.

Commenting on Krishnamurthy's analysis of the particular effectiveness of policy during times of distress, the state dependence as well as how conventional and unconventional policy work in different ways, Ludvigson remarked that it is very difficult to replicate accommodative monetary policy at the lower bound with QE. Pointing to the issue of channels to the broader economy, Ludvigson noted that unconventional policy may remove tail risk while conventional policy may push more on the means of the distribution. She wondered whether Krishnamurthy's interpretation of a Fed put was consistent with this.

Krishnamurthy responded by suggesting the Fed put is likely in play and that its actions in a tail state are being signaled through the Fed's announcements.

Brookings Papers on Economic Activity, Fall 2022: 276–280

Wright remarked that event studies have taught us a lot, but if we believe QE and QT are different—QE removes tail risk while QT may not—one should differentiate between announcements related to QE and those related to QT. On Ludvigson's results, Wright thought that the tapering announcements were not directly about balance sheet shrinkage but rather announcements about how QE was not going to last indefinitely. Wright concluded by saying that event study evidence that does not borrow the inference that QT is the opposite of QE would be useful but admitted that this is hard to do.

Michael Kiley disagreed with Krishnamurthy's characterization of the views of central bankers, and especially of Federal Reserve analyses, in which Krishnamurthy argues such institutions' policy approach is misguided and based on a view that QE and policy interest rate changes are very close substitutes. Kiley offered three supporting points. First, research published by the Federal Reserve System staff suggest that QE is not a perfect substitute for the federal funds rate; for example, research on the broad versus narrow channels of QE is dominated by Fed researchers and often emphasizes the narrow channel.[1] Kiley explained that this work also informs how simulations are run in—for example, the FRB/US model—and emphasized that the notion that QE is not a perfect substitute for the federal funds rate is often incorporated in such simulations.[2] Similarly, Kiley continued, research using a dynamic stochastic equilibrium model, as in early work by Edward Nelson and David Lopez-Salido among others, also directly incorporates this imperfect substitutability.[3] He summarized by saying that he does not agree with Krishnamurthy's analysis that the prevalent policy approach is somehow fundamentally flawed but he admitted that we have much more to learn on what good QE rules are.

1. For example, Stefania D'Amico and Thomas B. King, "Flow and Stock Effects of Large-Scale Treasury Purchases: Evidence on the Importance of Local Supply," *Journal of Financial Economics* 108, no. 2 (2013): 425–48.

2. For example, Michael T. Kiley, "Quantitative Easing and the 'New Normal' in Monetary Policy," *The Manchester School* 86, no. S1 (2018): 21–49, https://doi.org/10.1111/manc.12238.

3. For example, Javier Andrés, J. David López-Salido, and Edward Nelson, "Tobin's Imperfect Asset Substitution in Optimizing General Equilibrium," *Journal of Money, Credit and Banking* 36, no. 4 (2004): 665–90, http://www.jstor.org/stable/3839037; and Michael T. Kiley, "The Aggregate Demand Effects of Short- and Long-Term Interest Rates," *International Journal of Central Banking* 10, no. 4 (2014): 69–104.

Robert Hall distinguished between two parts in the discussion. Hall pointed out that QE is used for the Fed to put out fires in particular asset markets—the emergency move is always to purchase bonds, not to sell them, giving rise to the asymmetry we see in the literature. Regarding the debate on the optimal size of the Federal Reserve balance sheet, Hall noted that he belongs to the school that does not necessarily think the balance sheet needs to shrink. He then commented that the Fed is very active in the repo market, which is an important tool that does not get much attention. Hall concluded by arguing that the most important issue is coordination with the Treasury in terms of the maturity of reserves, stating that while the Fed prefers the federal debt to be funded in the overnight market, the Treasury likes to borrow long.

Jason Furman wondered whether the analysis by the panelists had any relevance for macroeconomic questions, including unemployment and inflation, and noted that he had interpreted the panelists response as a tentative "no."

Kristin Forbes asked the panelists if they had any thoughts on how other countries' monetary policies may affect the United States and noted that the situation today is very different than in 2017 because multiple countries around the world are now unwinding their balance sheets simultaneously.

Jonathan Pingle wondered whether the panelists were concerned about the speed with which the balance sheet is shrinking, substantially compressing the time between warning signs and stress.

Donald Kohn said that the 2013 taper tantrum was also primarily the result of the markets interpreting impending QT as a monetary policy signal, and the experience unfortunately contributed to a lot of inertia in the Fed's more recent response—when, how fast, and how to announce tapering in 2021. Kohn stated that there are other ways of signaling future monetary policy, however, which are more actively used today and allow the Fed to worry less about a taper tantrum now than would have been true in 2013. The equilibrium level of reserves and the volatility getting there depend importantly on the regulatory environment, he continued, noting that the demand for reserves depends partly on the liquidity regulations, and the Fed can do a lot to lower the demand for reserves by carefully structuring its financial stability regulations. Kohn argued that the leverage ratio is constraining dealers and that more needs to be done to make the Treasury market more liquid.

Joe Beaulieu was puzzled by what he interpreted as the ongoing angst about the QT ramp-up, even though it seems to be a done deal.

Jonathan Parker suggested that conventional policy also seems to have disproportionate effects. To the extent that QE does have unusual power, it is presumably because the market has the infrastructure and volume to manage policy-induced changes in short-term interest rates and their effects propagate smoothly across markets. Parker stated that what makes QE special is not the particular asset being purchased but the infrequency of its implementation and the policies and information that accompany the announcement of asset purchases. He argued that the Fed could likely steer the economy largely by intervening between reserves and long-term assets, rather than short-term assets. He summarized, stating that what is particular about QE is that it is infrequent and seems to act in the market more as if it were a surprise. Thus, the effects of QT may be quite different because it is a slow and expected policy which allows markets to be better prepared. Commenting on Ludvigson's results on the price response to inflexibility, Parker thought it would be interesting to distinguish between two different types of inflexibility: inflexible discretionary policy and rule-based inflexibility.

Hanno Lustig addressed Wright's finding of a small effect of QT—25 basis points—saying that he struggled to reconcile this result with recent events in the bond market. He provided the example of the ten-year TIPS (Treasury inflation-protected securities) yield and stated that the real yields have gone up by about 150 basis points in a matter of months. He concluded that while the Fed has started conventional monetary policy, this should have negligible effects ten years from now, leaving open the question of what may account for this dramatic move.

Benjamin Friedman had a general observation, saying that comments suggesting that quantitative actions do not have a proper theory puzzles him. Friedman offered the standard Markowitz and Tobin asset pricing models as examples of existing frameworks. He then argued that central banks should be prepared to use QT deliberately in specific circumstances. Leading up to the Great Recession with an overheated housing market, the standard story is that the interest rate was too blunt an instrument to attack overheating in one sector of the economy. But, Friedman contended, if the Fed had deliberately sold a substantial amount of its mortgage-backed securities, targeting the housing market specifically, the instrument would not have been blunt at all. Friedman suggested that the Fed keep a healthy supply of certain assets on its balance sheet to be able to implement QT in a targeted way when necessary.

Annette Vissing-Jorgensen first addressed the question of the scope of possible QT and noted that she had reestimated the amount of feasible runoff with results that are quite a bit higher than those presented by Wright.[4] Second, Vissing-Jorgensen pondered the extent to which QT should be used actively or passively. One possibility would be to use QT simply to shrink the balance sheet in preparation for the next downturn, opening up the possibility for more QE; the other possibility is to use it actively in a manner similar to the short rate. Vissing-Jorgensen noted that there has not been a lot of research or debate on the effectiveness of QT and called for more work on the issue.

4. David Lopez-Salido and Annette Vissing-Jorgensen, "Reserve Demand and Quantitative Tightening," working paper (2022), https://drive.google.com/file/d/161MMKn4pqRJTAY26ZaCF_1cMOYFpwOVy/view.

CEVAT GIRAY AKSOY
European Bank for Reconstruction and Development and King's College London

STEVEN J. DAVIS
University of Chicago Booth School of Business and Hoover Institution

JOSE MARIA BARRERO
Instituto Tecnológico Autónomo de México

MATHIAS DOLLS
ifo Institute

NICHOLAS BLOOM
Stanford University

PABLO ZARATE
Princeton University

Working from Home Around the World

ABSTRACT The pandemic triggered a large, lasting shift to work from home (WFH). To study this shift, we survey full-time workers who finished primary school in twenty-seven countries as of mid-2021 and early 2022. Our cross-country comparisons control for age, gender, education, and industry and treat the United States mean as the baseline. We find, first, that WFH averages 1.5 days per week in our sample, ranging widely across countries. Second, employers plan an average of 0.7 WFH days per week after the pandemic, but workers want 1.7 days. Third, employees value the option to WFH two to three days per week at 5 percent of pay, on average, with higher valuations for women, people with children, and those with longer commutes. Fourth, most employees were favorably surprised by their WFH productivity during the pandemic. Fifth, looking across individuals, employer plans for WFH levels after the pandemic rise strongly with WFH productivity surprises during the pandemic. Sixth, looking across countries, planned WFH levels rise with the cumulative stringency of government-mandated lockdowns during the pandemic. We draw on these results to explain the big shift to WFH and to consider some implications for workers, organization, cities, and the pace of innovation.

Conflict of Interest Disclosure: The authors' Global Survey of Working Arrangements (G-SWA) was supported by the European Bank for Reconstruction and Development, the Asociación Mexicana de Cultura AC, the Becker Friedman Institute at the University of Chicago, the ifo Institute, King's College London, the Smith Richardson Foundation, and the Templeton Foundation. 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. They 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, Fall 2022: 281–330 © 2023 The Brookings Institution.

The COVID-19 pandemic triggered a huge, sudden uptake in working from home, as individuals and organizations responded to contagion fears and government restrictions on commercial and social activities. Over time, it has become evident that the big shift to work from home (WFH) will endure after the pandemic ends. No other episode in modern history involves such a pronounced and widespread shift in working arrangements in such a compressed time frame. The shift from farms and craft production to factory jobs that accompanied the Industrial Revolution played out over roughly two centuries. The later, ongoing shift from factory work and other goods production to services is many decades in the making. While these transitions brought greater changes in skill requirements and business operations, their comparatively slow unfolding afforded much more scope for gradual adjustment.

These observations prompt some questions: What explains the pandemic's role as catalyst for a lasting uptake in WFH? What does a large, lasting shift to remote work portend for workers? Specifically, how much do they like or dislike WFH? How do preferences in this regard differ between men and women and with the presence of children? How, if at all, do workers and employers act on preferences over working arrangements? When looking across countries and regions, have differences in pandemic severity and the stringency of government lockdowns had lasting effects on WFH levels? Finally, how might the big shift to remote work affect the pace of innovation and the fortunes of cities?

To tackle these and related questions, we field a new Global Survey of Working Arrangements (G-SWA) in twenty-seven countries. The survey yields individual-level data on demographics, earnings, current WFH levels, employer plans, and worker desires regarding WFH after the pandemic, perceptions related to WFH, commute times, willingness to pay for the option to WFH, and more. Thus far, we have fielded the survey online in two waves, one in late July/early August 2021 and one in late January/early February 2022. Our G-SWA samples skew to relatively well-educated persons within each country, less so in most rich countries but very strongly so in middle-income countries.[1]

We focus our analysis on full-time workers, age 20–59, who finished primary school and investigate how outcomes, plans, desires, and perceptions around WFH vary across persons and countries. In making comparisons

1. This pattern is typical in online surveys covering many countries. See Alsan and others (2020), Stantcheva (2021), and Dechezleprêtre and others (2022).

across countries, we consider conditional mean outcomes that control for gender, age, education, and industry at the individual level, treating the raw US mean as the baseline value. These values should not be understood as estimated means for the working-age population or overall workforce in each country. Rather, they are conditional sample means for relatively well-educated, full-time workers who have enough facility with smartphones, computers, tablets, and the like to take an online survey.

Conditional mean WFH values average 1.5 full paid days a week across the countries in our sample as of mid-2021 and early 2022, ranging from 0.5 days in South Korea and 0.8 in Taiwan to 1.6 in the United States, 2.0 in the United Kingdom, and 2.6 in India. We also find that employers plan an average of 0.7 WFH days per week after the pandemic, but workers want 1.7 days, considerably more. Separate US data from the Survey of Working Arrangements and Attitudes (SWAA) also show a large gap between employer plans and worker desires in this regard.[2]

There are several reasons to think that WFH levels will ultimately settle at higher values than suggested by our survey data (for the well-educated groups covered by the G-SWA). SWAA updates at the WFH Research website show a steady rise from January 2021 to June 2022 in the plans of American employers for WFH levels after the pandemic. Similarly, G-SWA data show upward revisions over time in planned WFH levels for ten of the twelve countries covered by both survey waves. This pattern suggests that employers are gradually warming to the practice of letting employees work remotely one or two days per week in many jobs and most or all of the time in some jobs. Drawing on a near-universe of online job vacancy postings in the United States and four other English-speaking countries, Hansen and others (2022) find strong upward trajectories from mid-2020 through mid-2022 in the share of new vacancy postings that say employees can work remotely one or more days per week. Adrjan and others (2021) find the same pattern through September 2021 in vacancy postings for twenty OECD countries. This pattern suggests that remote work practices are becoming more firmly rooted, even as COVID-19 deaths decline. Finally, the share of US patent applications that advance video conferencing and other remote-interaction technologies doubled in the wake of the pandemic (Bloom, Davis, and Zhestkova 2021). This redirection of innovation efforts suggests that remote work technologies will continue to improve, further encouraging the use of remote work practices.

2. WFH Research, "Working from Home before and since the Start of COVID," www.WFHresearch.com.

How did the pandemic catalyze a large, lasting shift to WFH? We find strong evidence for a three-part explanation. First, the pandemic compelled a mass social experiment in WFH. Second, that experimentation generated a tremendous flow of new information about WFH and greatly shifted perceptions about its practicality and effectiveness. The simultaneity of experimentation across suppliers, producers, customers, and commercial networks yielded experience and information that was hard to acquire before the pandemic. Third, in light of this new information and shift in perceptions, individuals and organizations re-optimized working arrangements and moved to a much greater reliance on WFH. Barrero, Bloom, and Davis (2021c) sketch a theory that formalizes this three-part explanation and find supporting evidence for the United States. We investigate how this explanation fares in our twenty-seven-country sample.

Fears of contracting COVID-19 and government-mandated lockdowns drove workers and employers to experiment at scale with WFH. Because the pandemic lingered and recurred, workers and organizations experimented intensively with WFH for many months. This much is obvious. Less apparent is how the experimentation influenced perceptions about WFH and whether any shift in perceptions had a lasting impact on working arrangements. In this regard, we find two key results: first, relative to their pre-pandemic expectations, most workers were surprised to the upside by their WFH productivity during the pandemic. That is, by their own assessments, they were more productive in WFH mode than they had anticipated. Only 13 percent of workers were surprised to the downside, and nearly a third found WFH to be about as productive as expected. Second, the extent of WFH that employers plan after the pandemic rises strongly (in the cross section) with employee assessments of WFH productivity surprises during the pandemic. This pattern holds in all twenty-seven countries in our sample. It indicates that large-scale experimentation with WFH permanently shifted views about the efficacy of remote work and, as a result, drove a major re-optimization of working arrangements.

We also investigate whether societal experiences during the pandemic had lasting effects on WFH levels. One aspect of societal experiences is the stringency and duration of government restrictions on commercial and social activity, which we summarize in a cumulative lockdown stringency (CLS) index. A second aspect is the severity of the pandemic itself, as summarized by cumulative COVID-19 deaths per capita. In this regard as well, we find two key results. First, employers plan higher post-pandemic WFH levels in countries with higher CLS values in regression models that control for worker characteristics, survey wave, cumulative COVID-19 deaths, and log real GDP per capita. Raising the country-level CLS value by two standard deviations raises employer plans for the post-pandemic WFH level by an extra 0.27 days per week, according to the model. This effect is 38 percent as large as the cross-country mean of 0.7 planned WFH days per week. Second, and to our surprise, cumulative COVID-19 deaths per capita have no discernable impact on planned WFH levels (or actual WFH levels as of the survey).

The pandemic spurred several other developments that helped drive a large, lasting uptake in WFH: new investments in the home and inside organizations that facilitate WFH, learning by doing in the WFH mode (as distinct from learning by experimentation), advances in products and technologies that support WFH, much greater social acceptance of WFH, and lingering concerns about infection risks that lead some people to prefer remote work. The rise of the internet, the emergence of the cloud, and advances in two-way video technologies before the pandemic created the conditions that made possible a big shift to WFH. Indeed, the extent of remote work was trending slowly upward, from a low base, long before the pandemic.[3]

What does a large, lasting shift to remote work portend for workers? According to G-SWA data, employees view the option to WFH two to three days per week as equal in value to 5 percent of earnings, on average. The conditional mean willingness to pay for this option is positive for every country except Taiwan. Other survey responses tell a consistent story. For example, when we query respondents about how much they want to WFH after the pandemic, country-level conditional means range from 1.1 to 2.3 days per week. When we ask those who currently WFH one or more days per week how they would respond "if your employer announced that all employees must return to the worksite 5+ days a week," one-quarter say they would quit or seek a job that lets them WFH one or two days per week. Savings in commute time are perhaps the most obvious and important individual-level benefit of WFH. Daily round-trip commutes average 64 minutes per day in the G-SWA sample, ranging from 48 minutes in the United States and Serbia to 93 minutes in India and 96 minutes in China.

Women place a higher average value on WFH than men in all but a few countries, as do those with more education. Among married persons, both

3. Barrero, Bloom, and Davis (2022c, slide 6) draw on the American Time Use Survey and the American Community Survey to present evidence that the share of full paid days worked from home rose from 0.4 percent in 1965 to 1.0 percent in 1990, 2.8 percent in 2010, and 4.7 percent in 2019. Our discussant, Katharine Abraham, also presents evidence of an upward drift in US WFH rates from 1997 to 2018 based on data from the Survey of Income and Program Participation and the American Time Use Survey.

men and women more highly value the option to WFH when they have children under age 14. Not surprisingly, willingness to pay for WFH rises with commute time. All of these patterns emerge clearly in the data, but the heterogeneity in willingness to pay for WFH is perhaps even more noteworthy. Even when we control for education, age, gender, marital status, presence of children, commute time, current WFH days, survey wave, and country, the residual variation in willingness to pay is large, and our regression R^2 values are less than 12 percent. This preference heterogeneity has important implications for organizations and for policy, as we discuss.

We also offer several observations about how the rise of remote work could affect the pace of innovation and the fortunes of cities. With respect to innovation, we argue that there are sound reasons for optimism. With respect to cities, we highlight some major challenges—especially for urban centers that, before the pandemic, organized themselves to support high-volume inward commuting and a high spatial concentration of commercial activity. A key point is that the rise of remote work raises the sensitivity of the city-level tax base with respect to the quality of its governance and local amenities. For poorly governed cities, in particular, this greater sensitivity raises the risk of a downward spiral in local tax revenues and urban amenities.

Our study relates to many previous works. We build on the US-centric analysis of Barrero, Bloom, and Davis (2021c) and borrow heavily from their SWAA questionnaire in designing our survey questions. Criscuolo and others (2021) survey managers and employees about their experiences and expectations around WFH in twenty-five countries. They find "a large majority of managers and workers had a positive experience from teleworking" (7) during the pandemic, which aligns well with our evidence and with evidence for American managers and workers in Ozimek (2020) and Barrero, Bloom, and Davis (2021c). Criscuolo and others (2021) also investigate how managerial experiences relate to future WFH levels in their organizations. Managers who more favorably assess their company's experience with telework during the COVID-19 crisis prefer higher WFH levels for their company in the future, even when controlling for the extent of WFH at the company before and during the pandemic. Their evidence from a survey of managers covering many countries strongly aligns with our evidence from a survey of workers.

Many studies examine the huge uptake in WFH in spring 2020.[4] Our surveys went to field 16 to 23 months after the pandemic's onset and reflect

4. See, for example, Adams-Prassl and others (2020), Barrero, Bloom, and Davis (2020), Bartik and others (2020), Bick, Blandin, and Mertens (2020), Brynjolfsson and others (2020), Eurofound (2020), and Ker, Montagnier, and Spiezia (2021).

experiences and perceptions at that time. Previous studies also document preference heterogeneity around WFH in various settings and using a range of empirical methods.[5] Relative to these studies, we contribute by documenting the pervasiveness of heterogeneity in WFH preferences around the world and by showing that the structure of preferences exhibits common features across countries, including stronger desires to WFH among those with children. Other studies stress the economic resilience value of WFH during a pandemic and its role in slowing the spread of the SARS-CoV-2 virus.[6]

Adrjan and others (2021) find that differences across countries in government lockdowns during the pandemic and "digital preparedness" before the pandemic partly explain cross-country differences in the persistent shift to remote work. Baker, Davis, and Levy (2022) find that government lockdown stringency during the pandemic had persistent effects on state-level unemployment rates in the United States. These results align with our evidence that societal experiences during the pandemic have persistent effects on the extent of WFH. Our concerns about how remote work presents challenges for cities, especially poorly governed ones, overlap with concerns expressed in Glaeser (2022).

I. The Global Survey of Working Arrangements (G-SWA)

The G-SWA covers full-time workers, age 20–59, who finished primary school in twenty-seven countries.[7] In addition to basic questions on demographics, employment status, earnings, industry, occupation, marital status, and living arrangements, the survey asks about current, planned, and desired WFH levels, perceptions and experiences related to WFH, willingness to pay for the option to WFH, commute time, and more. We design the G-SWA instrument, adapting questions from the US SWAA developed by Barrero, Bloom, and Davis (2021c). We enlist professionals to translate our original English-language questionnaire into the major languages of each country.[8] To ensure high-quality translations, we also enlist an independent

5. See, for example, Bloom and others (2015), Mas and Pallais (2017), Wiswall and Zafar (2018), He, Neumark, and Weng (2021), Barrero, Bloom, and Davis (2021c), and Lewandowski, Lipowska, and Smoter (2022).

6. See Alipour, Fadinger, and Schymik (2021), Bai and others (2021), Berniell and others (2021), Barrero, Bloom, and Davis (2021b), and Eberly, Haskel, and Mizen (2021).

7. Wave 1 includes part-time workers and those who did not finish primary school, but we omit them in our analysis.

8. The G-SWA survey instruments are available at https://wfhresearch.com/wp-content/uploads/2022/07/G-SWA_Wave1.pdf and https://wfhresearch.com/wp-content/uploads/2022/07/G-SWA_Wave2.pdf.

third party with knowledge of the survey to review the translations and revise as needed.

To field the G-SWA, we contract with Respondi (a professional survey firm), which implements the survey directly and in cooperation with its external partners. The survey effort taps pre-recruited panels of people who previously expressed a willingness to take part in research.[9] Recruitment into these panels happens via partner affiliate networks, multiple advertising channels (including Facebook, Google AdWords, and other websites), address databases, and referrals. New recruits are added to the panels on a regular basis. When it is time to field a survey, Respondi or its partner issues email messages that invite panel members to participate. The message contains information about compensation and estimated completion time but not about the survey topic. Clicking on the link in the invitation message takes the recipient to the online questionnaire. Respondents who complete the survey receive cash, vouchers. or award points, which they can also donate.[10]

This survey technology meets two market tests. First, it is increasingly used in scholarly research to examine preferences, attitudes, and perceptions and to field experiments. See Alesina, Stantcheva, and Teso (2018) for an early multicountry application. Second, reliance on pre-recruited samples for online surveys has exploded in market research studies and other commercial applications. We know of no comprehensive statistics on the scale of this activity, but consider Cint Group AB, a listed firm, that describes itself as "one of the world's largest consumer networks for digital survey-based research."[11] According to its website at the time of writing, Cint had 239 million or more engaged respondents across 130 countries, and it operated more than 4,600 survey panels tapped by more than 3,200 clients, including Zappi, SurveyMonkey, Qualtrics, Ipsos, and Nielsen. Commercial use on this scale suggests that sampling from pre-recruited panels to conduct online surveys can deliver useful insights in multiple domains and on many topics.

The G-SWA went to field in fifteen countries in late July and early August 2021 and in an overlapping set of twenty-five countries in late January and early February 2022. Wave 2, which covered both Russia and Ukraine, went to field shortly before the onset of the Russian invasion but well after Russia began massing troops near the Ukrainian border. We

9. Respondi and its external partners do not engage in "river sampling," whereby people are invited to take a survey while engaging in another online activity. Relative to river sampling, the use of pre-recruited panels affords greater control over sample composition and selection.

10. We do not contact respondents ourselves, do not collect personally identifiable information, and have no way to recontact them.

11. Cint, home page, https://www.cint.com.

retain the Ukrainian and Russian data in our study but acknowledge that war concerns may affect outcomes, attitudes, and perceptions related to WFH. Some G-SWA country waves include additional survey blocks that come after the demographic, employment, and WFH blocks.

Before proceeding to our analysis of the G-SWA data, we drop "speeders," defined as respondents in the bottom 5 percent of the completion-time distribution for each country. We also drop the roughly 15 percent of respondents who fail the following attention-check question: "In how many big cities with more than 500,000 inhabitants have you lived? . . . [T]his question only serves the purpose to check your attention. Irrespective of your answer, please insert the number 33." The resulting analysis sample contains 12,229 observations across fifteen countries in wave 1 and 23,849 observations across twenty-five countries in wave 2. Online appendix table A.1 reports observation counts and dates in the field for each country and survey wave. Tables A.2 and A.3 report summary statistics for key G-SWA variables. Median survey completion times range from 7.3 to 9.5 minutes, after drops, across the ten country waves that do not have extra survey blocks.

Although Respondi aims for samples that are broadly representative by age, gender, income, and regions within countries, our G-SWA samples are not representative of country-level workforces or their working-age populations. Respondents take the survey on a computer, smartphone, iPad, or like device, so we miss persons who don't use such devices. The G-SWA samples skew toward relatively well-educated persons in each country, less so in most advanced economies but very strongly so in some advanced economies and in middle-income economies. That could influence our results, even when we condition on certain observables.

Online appendix table A.4 compares our country-level G-SWA samples to Gallup data for 2017–2018. The comparisons suggest that our samples are reasonably representative of full-time workers, age 20–59, who finished primary school, with respect to age and gender, except for an overrepresentation of women in a few countries, especially India and Turkey. Most of our country-level samples are highly skewed to college-educated persons. In China, for example, 90 percent of G-SWA respondents completed college as compared to only 27 percent in the Gallup data.[12] Accordingly,

12. Gallup data have their own oddities, greatly underrepresenting college-educated persons in Spain, for example. In unreported results, we find that Gallup-based statistics for the share of persons age 25 and older with a college degree often differ by 10 percentage points or more (in both directions) from analogous statistics obtained from the World Bank and the European Social Survey. The World Bank and European Social Survey statistics also differ from each other, sometimes by 10 percentage points or more (again, in both directions).

when we report country-level (conditional) mean values, we use “HE” to designate countries with G-SWA samples that greatly overrepresent highly educated persons. When we investigate how societal experiences during the pandemic relate to post-pandemic outcomes, we consider the sensitivity of our results to samples that restrict attention to college-educated workers.

II. Working from Home in Twenty-Seven Countries

II.A. WFH Levels, Plans, and Desires

Figure 1 highlights the global nature of WFH among well-educated workers as of mid-2021 and early 2022. It reflects responses to the G-SWA question, “How many *full paid working days* are you *working from home* this week?” Response options range from none to 5 or more days per week.[13] The figure reports conditional mean responses, which we obtain from the coefficients on country-level dummies in an ordinary least squares (OLS) regression, treating the raw US mean as the baseline. The regression controls for gender, age groups (20–29, 30–39, 40–49, 50–59), education groups (secondary, tertiary, graduate), eighteen industry sectors, and survey wave. Online appendix A explains this conditioning method in fuller detail. Here and elsewhere, we include self-employed persons except when using data on employer plans. We pool over the mid-2021 and early 2022 survey waves when available and otherwise use data from a single wave.

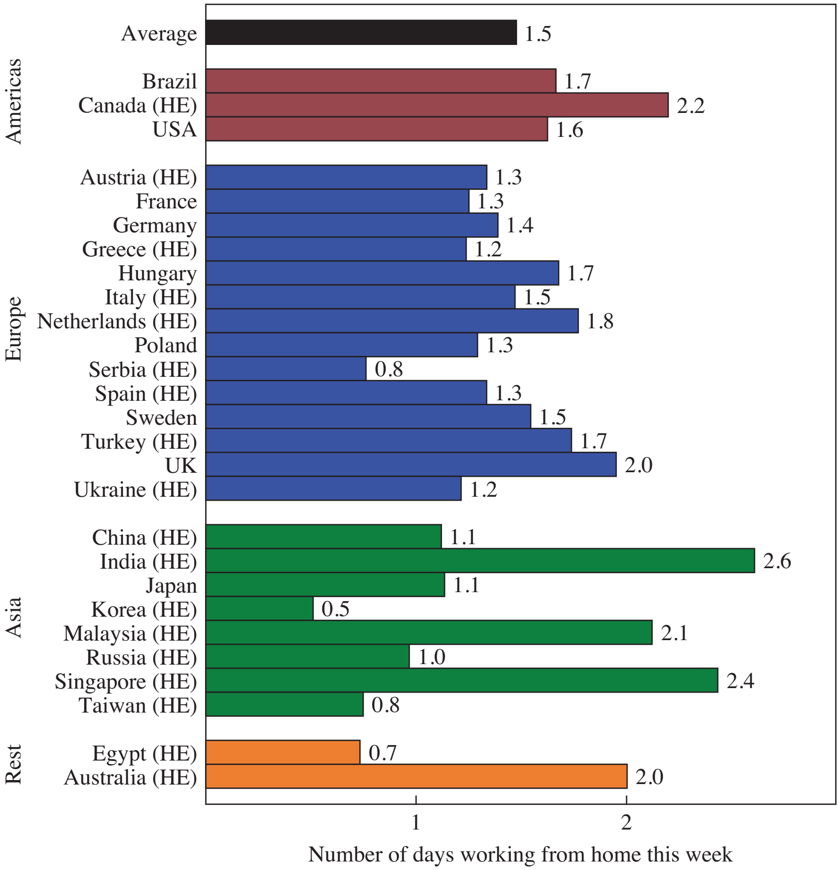
Full WFH days average 1.5 per week across the countries in our sample. We compute this average as the simple mean of the country-level conditional means. These conditional mean values range widely from 0.5 days in South Korea, 0.7 in Egypt, and 0.8 in Serbia and Taiwan at the low end to 2.4 in Singapore and 2.6 in India at the high end. The United States is in the middle at 1.6 WFH days per week. The wide dispersion in WFH levels conditional on individual characteristics, industry, and calendar time partly motivates our investigation into whether societal experiences during the pandemic had long-lasting effects on working arrangements.

Figure 2 provides direct evidence that high WFH levels will persist beyond the pandemic. The underlying question is “*After COVID, in 2022*

13. Katharine Abraham points out that our survey data could be affected by primacy bias, the tendency of respondents to pick answers that appear earlier in the list of response options. It’s a good point, and we plan to randomize the ordering of response options in future G-SWA waves. That said, our practice of dropping speeders will eliminate respondents who simply click on the first option. Our short survey instrument and the omission of persons who fail the attention-check question will mitigate any tendency to pick early options that arises from survey fatigue or inattentiveness.

Figure 1. Working from Home Is Now a Global Phenomenon among the Well-Educated

Paid full days working from home in the survey week, country-level conditional means

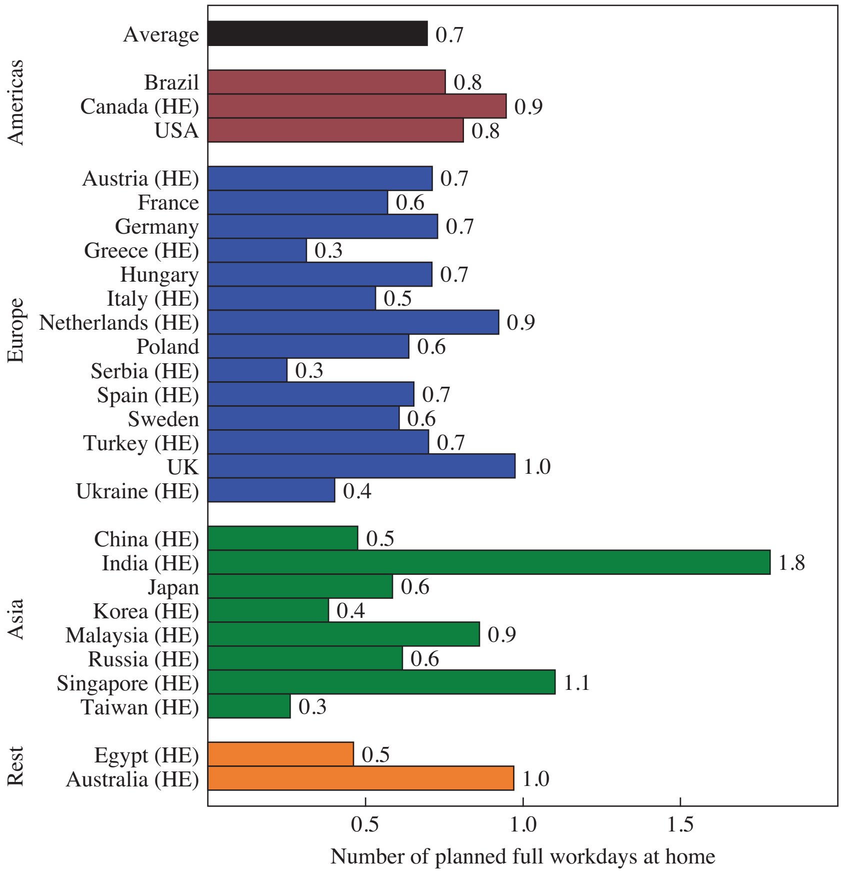


Source: Authors' calculations using G-SWA data.

Note: The survey question read: "How many *full paid working days* are you *working from home* this week?" The chart reports coefficients on country dummies in OLS regressions that control for gender, age (20–29, 30–39, 40–49, 50–59), education (secondary, tertiary, graduate), eighteen industry sectors, and survey wave, treating the raw US mean as the baseline value. We fit the regression to data for 33,091 G-SWA respondents surveyed in mid-2021 and early 2022. "Average" refers to the simple mean of the country-level values.

Figure 2. Planned Levels of Working from Home after the Pandemic

Average number of WFH days per week that employers plan



Source: Authors' calculations using G-SWA data.

Note: The survey question read: "*After COVID, in 2022 and later*, how often *is your employer planning* for you to work full days at home?" The chart reports coefficients on country dummies in OLS regressions that control for gender, age, education, industry, and survey wave, treating the raw US mean as the baseline value. We fit the regression to data for 34,875 G-SWA respondents who were surveyed in mid-2021 and early 2022. We limit the sample to persons with an employer in the survey week. "Average" refers to the simple mean of the country-level values.

and later, how often *is your employer planning* for you to work full days at home?" If the worker says his or her employer has neither discussed the matter nor announced a policy regarding WFH, we assign a zero value. Employers plan an average of 0.7 WFH days per week after the pandemic, ranging from 0.3 days in Greece, Serbia, and Taiwan to 0.4 in South Korea and Ukraine to 1.0 in Australia and the United Kingdom and 1.8 in India. The United States is again close to the middle at 0.8 planned WFH days per week. As in figure 1, there is a wide dispersion in the country-level conditional mean values.

When we ask workers how many full days per week they would like to WFH after the pandemic, we obtain even higher levels, as shown in figure A.1 in the online appendix. On average across countries, employees want 1.7 WFH days per week after the pandemic ends. The country-level conditional mean values for desired WFH days range from 1.1 in China, 1.2 in South Korea, and 1.3 in France and Taiwan at the low end to 2.2 in Canada and 2.3 in Brazil and Singapore at the high end. For the United States, mean desired WFH days are 2.1 per week.[14] Employees want more WFH days per week than employers plan in every country, and the gap exceeds half a day per week in all countries except India.

The gap between employee desires to WFH after the pandemic and employer plans is also a striking feature of the separate SWAA data for the United States (Barrero, Bloom, and Davis 2021c). The SWAA tracks desires and plans in this regard at a monthly frequency and shows a steady fall from a peak gap of 1.4 days per week in December 2020 to 0.6 days in June 2022.[15] Upward revisions in employer plans account for 69 percent of this shrinking gap.

When we look at planned WFH levels in countries covered by both G-SWA waves, we find that ten of twelve experienced an upward revision in their conditional mean values over the six-month period from the mid-2021 wave to the early 2022 wave. The cross-country average increase over this period is 0.18 days per week. SWAA data for the United States show an upward revision of 0.57 days per week over the eleven-month period from

14. According to SWAA data, American workers desire an average 2.2 WFH days per week as of February 2022 (Barrero, Bloom, and Davis 2021c). According to Gallup's State of the Workforce survey in May/June 2021, 91 percent of American workers who worked at least some of their hours remotely hoped that they could continue to do so after the pandemic (Saad and Wigert 2021).

15. Monthly SWAA statistics for US WFH levels, plans, and desires are available at https://wfhresearch.com/wp-content/uploads/2022/07/WFHtimeseries_monthly.xlsx. The underlying micro data can be accessed at https://wfhresearch.com/data/.

July/August 2021 (timing of G-SWA wave 1) to June 2022 and 0.24 days per week over the five-month period from January/February 2022 (G-SWA wave 2) to June 2022. These observations indicate that figure 2 understates the levels to which WFH days per week will eventually settle.

II.B. People Like WFH

Figure 3 suggests that people highly value the opportunity to WFH. Indeed, when asked directly, G-SWA respondents say the option to WFH two to three days a week is worth 5 percent of earnings, on average. We elicit the willingness to pay for this option using a two-part question structure. First, we ask, "*After COVID, in 2022 and later,* how would you feel about working from home *2 or 3 days* a week?" If the response is "Neutral," we code the willingness to pay as zero. If the response is "Positive—I would view it as a benefit or extra pay," we follow up with "How much of a *pay raise* (as a percent of your current pay) would you value as much as the option to work from home 2 or 3 days a week?" There are six bucketed response options, ranging from "Less than a 5% pay raise" to "More than a 35% pay raise."[16] If the response is "Negative—I would view it as a cost or a pay cut," we follow up with a parallel question that replaces "*pay raise*" with "*pay cut*."

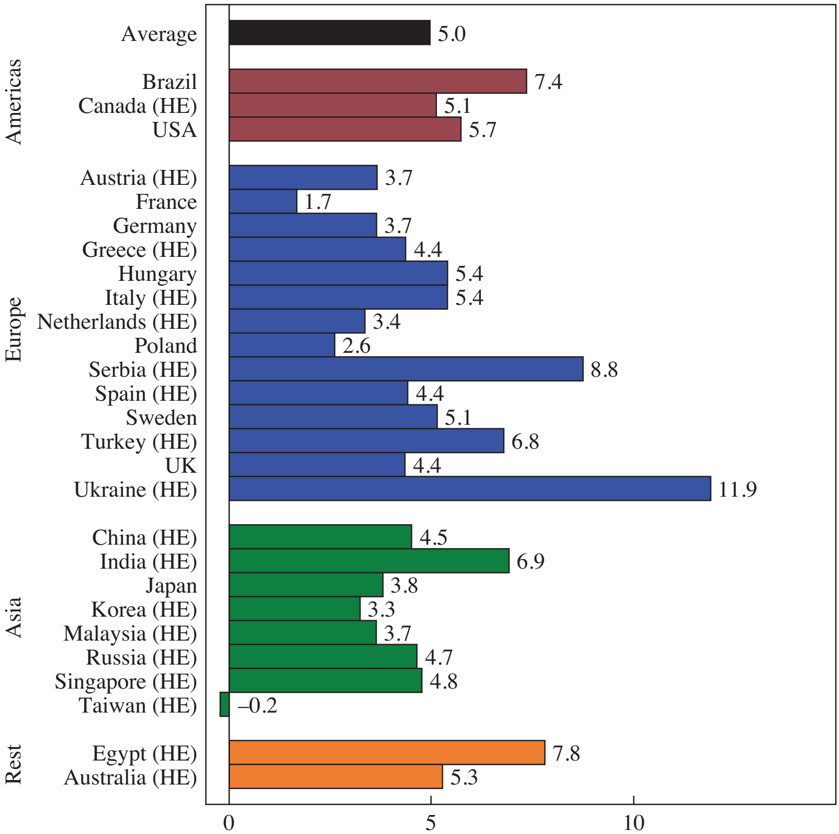
We use the two-part responses to quantify each person's willingness to pay and then construct the conditional mean values in figure 3. On average across countries, employees value this WFH option at 5 percent of pay. The country-level conditional mean willingness to pay is slightly negative for Taiwan and positive for all other countries, ranging upward to about 7–8 percent of pay in Brazil, Egypt, India, and Turkey, 8.8 percent in Serbia, and nearly 12 percent in Ukraine.

Other evidence reinforces the view that many employees like to WFH at least some of the time. The desired level of WFH averages 1.7 days per week across the countries in our sample (online appendix figure A.1). As shown in the online appendix, figure A.2, 26 percent of employees who currently WFH one or more days per week would quit or seek a job that allows WFH, if their employers were to require a return to five or more days per week on-site. Using SWAA data for US workers, Barrero, Bloom, and Davis (2021a) find that more than 40 percent of those who currently

16. The survey instrument includes both "A 25% to 35% pay raise" and "More than a 35% pay raise" options that we combine into one bucket for 25 percent or more. For persons in this top bucket, we assign a willingness-to-pay value of 25 percent. For the other buckets, we assign the midpoint value. We take the same approach for those who report a negative willingness to pay.

Figure 3. Willingness to Pay for the Option to Work from Home

Average amenity value of the option to work from home two to three days per week, as a percentage of pay



Source: Authors' calculations using G-SWA data.

Note: The survey questions read: "*After COVID-19, in 2022 and later*, how would you feel about working from home *2 or 3 days* a week?" and "How much of a pay raise [cut] (as a percent of your current pay) would you value as much as the option to work from home 2 or 3 days a week?" The chart reports coefficients on country dummies in OLS regressions that control for gender, age, education, industry, and survey wave, treating the raw US mean as the baseline value. We fit the regression to data for 36,078 G-SWA respondents who were surveyed in mid- 2021 and early 2022. "Average" refers to the simple mean of the country-level values.

WFH one or more days per week would quit or seek a new job if their employers were to require a full return to the company work site. Bloom and others (2015) designed a WFH field experiment for a large Chinese travel agency. When offered the option to WFH four days a week for nine months, with a fifth workday in the office, half the employees wanted to do so. Mas and Pallais (2017) integrate a field experiment into the application process for call center jobs by randomizing over combinations of pay and working arrangements. They use the resulting data to construct an implied willingness-to-pay distribution for the option to WFH, obtaining a mean value of 8 percent. Bloom, Han, and Liang (2022) conduct a randomized control trial of engineers and marketing and finance employees in a large technology firm, letting some of them WFH on Wednesday and Friday. This hybrid WFH arrangement cut quits by 35 percent and raised self-reported work satisfaction. After Spotify adopted a "work from anywhere" policy, attrition rates fell 15 percent in 2022:Q2 relative to 2019:Q2 (Kidwai 2022). This fall coincided with sharply increased quit rates for the overall economy.

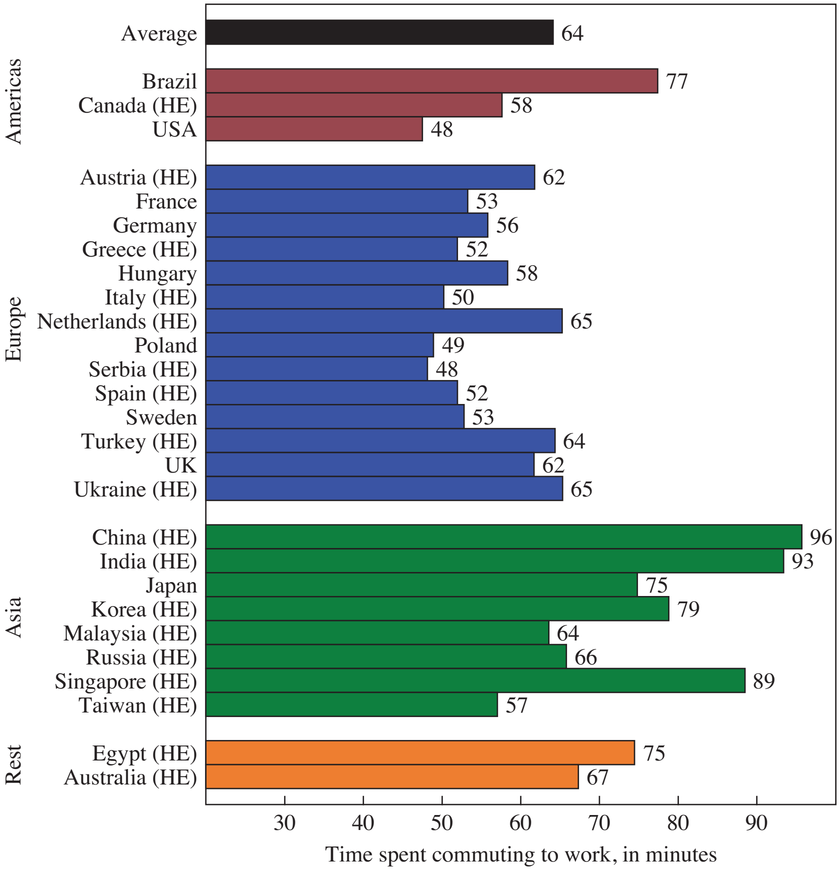
We see it as no surprise that (most) people place a sizable value on the option to WFH a few days per week. WFH saves on time and money costs of commuting. As shown in figure 4, round-trip commute times average 64 minutes per day in our sample, ranging from 48 minutes in Serbia and the United States to more than 90 minutes per day in China and India. WFH also economizes on grooming time and costs and affords more flexibility in time use over the day, greater personal autonomy, and less traffic-related stress.[17] Because the WFH amenity value is untaxed, it is more valuable for workers who face higher tax rates. The puzzle, if there is one, is why WFH levels were so low before the pandemic, given the now-evident practicality of much higher WFH levels than prevailed before March 2020.

Barrero, Bloom, and Davis (2021c) present survey evidence of what American workers like and dislike about WFH and about working on business premises. When asked "What are the top benefits of working from home?" and allowed to select up to three options, 51 percent say "No commute," 44 percent say "Flexible work schedule," 41 percent say "Less time getting ready for work," 37 percent say "Quiet," and 18 percent say "Fewer meetings." When asked "What are the top benefits of working on your employer's business premises?" 49 percent say "Face-to-face collaboration," 49 percent say "Socializing," 41 percent say "[Maintaining] work/personal

17. See, for example, Mas and Pallais (2017), Angelici and Profeta (2020), Barrero, Bloom, and Davis (2021a, 2021c), and Saad and Wigert (2021).

Figure 4. Commute Times Average More Than One Hour per Day

Daily round-trip commute time, in minutes



Source: Authors' calculations using G-SWA data.

Note: The survey questions read: Wave 1: "In 2019 (before COVID) how long was your typical commute to work in minutes (one-way)?" and Wave 2: "How long do you usually spend commuting to and from work (in minutes)? If you are not currently commuting to work, please answer based on your commute time in 2019 (before COVID)." The chart reports regression- adjusted conditional means, as in the previous figures. We fit the regression to data for 36,078 G-SWA respondents surveyed in mid-2021 and early 2022.

life boundaries," and 40 percent say "Better equipment." [18] Thus, both WFH and working on business premises have their attractions.

According to SWAA data from February to June 2022, most full-time American employees in jobs where remote work is feasible would like to split their workweeks between home and business, and most of the rest would like to WFH five days a week (Barrero, Bloom, and Davis 2022b, slide 22). Gallup's State of the Workforce survey conducted in May/June 2021 shows the same pattern (Saad and Wigert 2021). Barrero, Bloom, and Davis (2021c) quantify the time-saving gains for American workers from the pandemic-induced rise in WFH. Kahn (2022, chapters 2 and 3) offers an extended discussion of how WFH expands personal freedom, improves life quality, brings new employment opportunities, and builds social capital in residential communities.

II.C. The Structure of Preferences over WFH

Table 1 explores the structure of preferences around the option to WFH two to three days per week. We regress the willingness to pay for this option on individual characteristics, marital status, the presence of children, and commuting time. Several patterns emerge. Women more highly value the option to WFH than men, with an estimated differential that exceeds 1 percent of pay. People living with children under age 14 more highly value WFH, again with a differential greater than 1 percent of pay. Married women more highly value the option to WFH than single women, but the differential is modest. Not surprisingly, the WFH amenity value rises with commute time. The willingness to pay for the option to WFH also rises strongly with education. Column 3 says that graduate degree holders value the option to WFH at an extra 2.5 percent of pay relative to those with a secondary education. At least in part, this pattern probably reflects more spacious and comfortable homes and better internet quality among the more educated, in line with evidence for the United States in Barrero, Bloom, and Davis (2021b, 2021c).

When we expand the table 1 specifications to include flexible controls for the respondent's current WFH days per week, the education effect on willingness to pay shrinks by roughly a third and the R^2 values rise by about 3 percentage points. Otherwise, the same patterns continue to hold. Adding a control for self-assessed propensity to social distance and replacing coarse age bins with two-year age bins has little impact, except to improve

18. See slide 27 in Barrero, Bloom, and Davis (2022b), which tabulates SWAA data from February through June 2022.

Table 1. The Structure of Preferences over Working from Home

	Dependent variable: Amenity value of option to work from home two to three days a week				
	(1)	*(2)*	*(3)*	*(4)*	*(5)*
Tertiary education	1.19***	1.06***	1.23***	1.31***	1.17***
	(0.38)	(0.37)	(0.21)	(0.24)	(0.28)
Graduate degree	3.17***	3.02***	2.47***	2.78***	2.12***
	(0.24)	(0.23)	(0.35)	(0.46)	(0.38)
Married	0.34	0.34	0.36	0.23	0.51**
	(0.22)	(0.23)	(0.21)	(0.32)	(0.21)
1 (men)	−1.11***	−1.14***	−1.17***		
	(0.22)	(0.23)	(0.25)		
1 (lives with children under 14)	1.27***	1.21***	0.92***	1.07***	0.72**
	(0.33)	(0.32)	(0.30)	(0.29)	(0.27)
1 (men) times 1 (lives with children under 14)	0.06	0.06	0.005		
	(0.50)	(0.50)	(0.48)		
Round-trip commute time in hours		0.68***	0.66***	0.60***	0.72***
		(0.19)	(0.15)	(0.11)	(0.22)
Sample	All	All	All	Men	Women
Dependent variable SD	11.293	11.293	11.293	11.313	11.234
Observations	26,689	26,689	26,689	13,605	13,084
R^2	0.035	0.039	0.074	0.070	0.078
Country FE			Y	Y	Y

Source: Authors' calculations using G-SWA data.

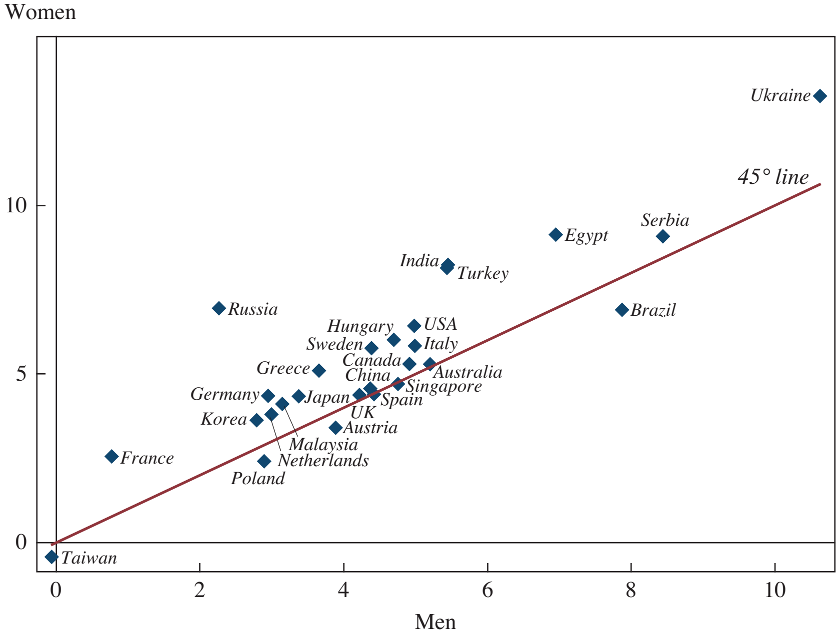
Note: The dependent variable is the willingness to pay for the option to work from home two to three days per week, computed using the two-part question structure described in the main text. The sample contains individual-level data in the twenty countries for which we have data on the number of children and marital status. All specifications include fixed effects for age groups and survey wave. We cluster errors at the country level.

$^{**}p < .05$, $^{***}p < .01$.

fit. In a more flexible nonparametric specification, the willingness to pay to WFH two to three days per week exceeds 2 percent of pay for someone with a round-trip commute of more than one hour relative to an observationally similar person who commutes less than 20 minutes per day.

Figures 5 and 6 provide evidence on how the structure of preferences around WFH varies across countries. We construct these figures using the same data and specifications as in figure 3, except we now fit the regressions separately for each subsample, for example, men and women. Figure 5 shows that women more highly value the option to WFH in most countries. The same pattern holds when we constrain the covariate coefficients to be the same for women and men, as suggested by the similarity of coefficients in columns 4 and 5 of table 1. The same pattern also holds when we

Figure 5. Women More Highly Value the Option to Work from Home in Most Countries



Source: Authors' calculations using G-SWA data.

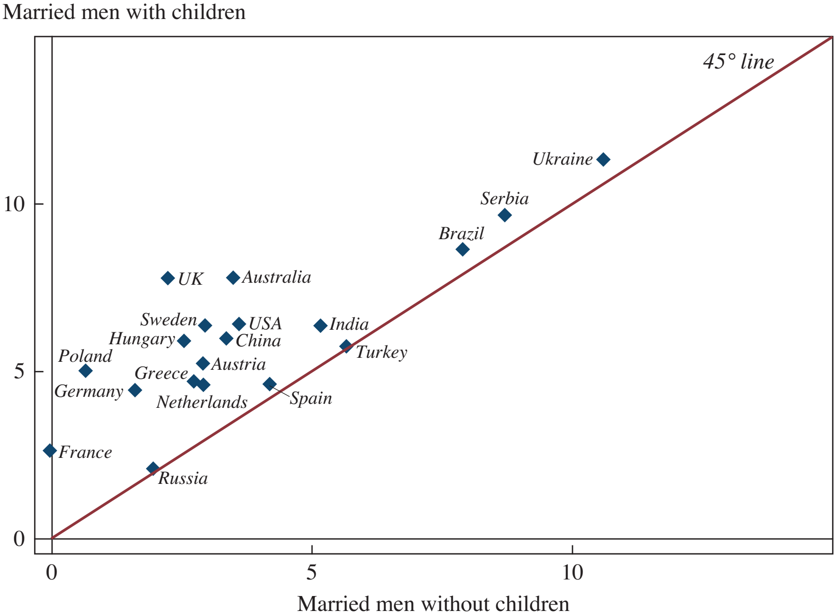
Note: This figure draws on the same questions and data as figure 4. It also uses the same specification, except that we fit the regression separately for men and women.

restrict attention to single persons with no children, as shown in panel C of figure 6. Thus, there appears to be a widespread pattern whereby women place more value on the option to WFH than men. It also appears that child-care responsibilities do not explain this pattern since we control for the presence of children, and the pattern also holds when we compare single women to single men. It may be that women, more than men, take on other caregiving and household management responsibilities that lead them to place a higher value on the flexibility and time savings afforded by the option to WFH.

Panels A and B in figure 6 highlight another commonality in the structure of preferences across countries: both men and women place a higher premium on the option to WFH when there are children in the household. We see this pattern as indicative of greater time demands and greater complexity in household management for people with children. As a result, they place greater value on the time savings and flexibility afforded by the option to WFH.

Figure 6. How the Amenity Value of Working from Home Differs by Sex and Family Circumstances, Conditional Means by Country

Panel A: Married men, comparison between those with and without children



Panel B: Married women, comparison between those with and without children

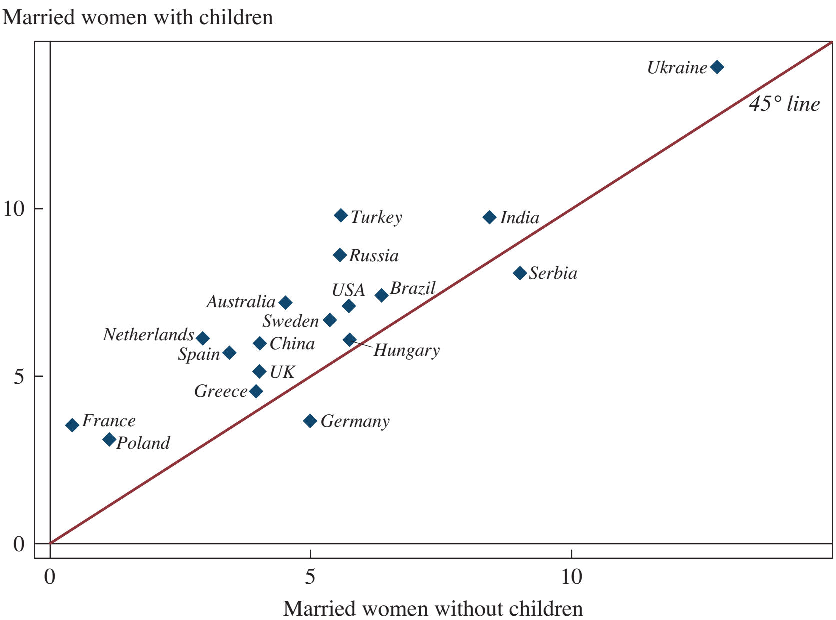
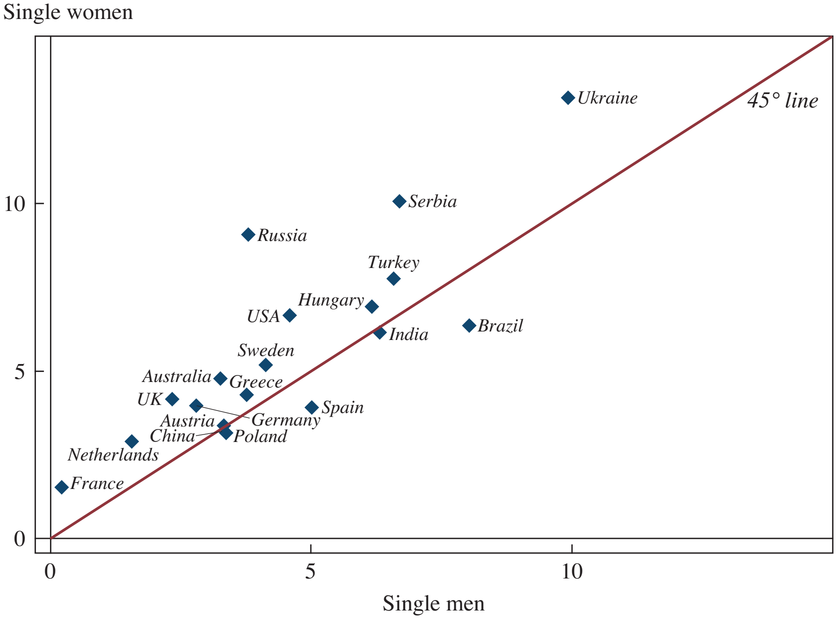


Figure 6. How the Amenity Value of Working from Home Differs by Sex and Family Circumstances, Conditional Means by Country (*Continued*)

Panel C: Unpartnered or single persons, comparison between men and women



Source: Authors' calculations using G-SWA data.
Note: The regression specification is the same as in figures 4 and 5, but we fit six separate regressions, one for each of the indicated subsamples. The charts suppress values for countries with fewer than fifty observations in the relevant sample (Egypt in all three panels, and Austria in panel B).

We make two additional observations about these results. First, table 1 and figures 5 and 6 imply large mean differences in the willingness to pay between well-defined groups. Consider two hypothetical persons: a married woman with a graduate degree, children under age 14, and a 45-minute one-way commute from her suburban home; and a single, college-educated man who lives 5 minutes from the office. This hypothetical woman values the WFH option at an extra 4.6 percent of pay compared to the hypothetical man, according to column 3 of table 1. The differential is 5.8 percent of pay with a nonparametric specification for commute time in an otherwise identical regression. We could construct comparisons that yield larger differences by considering worker age, for example. If table 1 and figures 5 and 6 provide a reasonably accurate portrayal of preferences, workers will (happily) sort across WFH levels that differ systematically between men and women, people with and without children, commuting time, and more.

Second, although the G-SWA data exhibit strong regularities in the structure of preferences around WFH, none of our statistical models account for a large share of willingness-to-pay variation. Even when we expand the table 1 specifications to include controls for current WFH days, replace coarse age bins with two-year bins, and relax linearity over commute time, the R^2 values never reach 0.12. While measurement error may play a role here, we see the modest R^2 values as an important result. Along with the dispersed response distribution for the dependent variable (online appendix figure A.3), the modest goodness of fit in these regressions says that people differ greatly in how much they value WFH. Moreover, readily observable attributes of persons account for only a modest share of this heterogeneity.

III. How the Pandemic Catalyzed a Big Shift to WFH

III.A. Pandemic-Induced Experimentation and Re-optimization of Working Arrangements

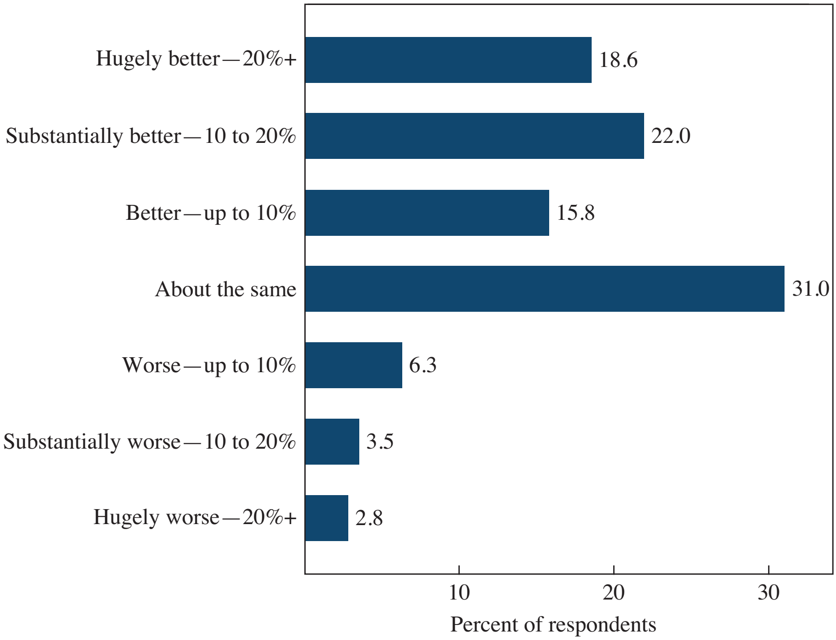
To explore the impact of pandemic-induced experimentation on perceptions about WFH productivity, we put the following question to G-SWA participants who mainly worked from home at some point during the pandemic: "Compared to your expectations *before COVID (in 2019)* how has working from home turned out for you?" Response options are as follows:

a) Hugely better—I am 20%+ more productive than I expected
b) Substantially better—I am 10% to 20% more productive than I expected
c) Better—I am 1% to 10% more productive than I expected
d) About the same
e) Worse—I am 1% to 10% less productive than I expected
f) Substantially worse—I am 10% to 20% less productive than I expected
g) Hugely worse—I am 20%+ less productive than I expected

Figure 7 shows the raw response distribution in the pooled G-SWA data.

This response distribution has two important features. First, it is highly dispersed. Since WFH levels were quite low before the pandemic—about 0.25 full days per week, according to the American Time Use Survey—wide dispersion in productivity surprises leads to persistently higher WFH levels. To see the logic, suppose for the moment that employer assessments of WFH productivity surprises align with employee assessments,

Figure 7. The Distribution of WFH Productivity Relative to Expectations



Source: Authors' calculations using G-SWA data.

Note: The survey question read: "Compared to your expectations *before COVID (in 2019)* how has working from home turned out for you?

a) Hugely better—I am 20%+ more productive than I expected
b) Substantially better—I am 10% to 20% more productive than I expected
c) Better—I am 1% to 10% more productive than I expected
d) About the same
e) Worse—I am 1% to 10% less productive than I expected
f) Substantially worse—I am 10% to 20% less productive than I expected
g) Hugely worse—I am 20%+ less productive than I expected"

The sample consists of 19,027 G-SWA respondents in mid-2021 and early 2022 who worked mainly from home at some point during the COVID-19 pandemic.

and consider the effects of dispersed WFH productivity surprises. For ease of exposition, assume for now that the willingness to pay to WFH is zero. In jobs and tasks perceived before the pandemic to be marginally less productive when performed remotely, positive WFH productivity surprises trigger a lasting shift to WFH mode. In contrast, zero and negative WFH productivity surprises lead to no re-optimization in jobs and tasks that were already perceived to be less productive in remote mode. Thus, given the low WFH levels that prevailed before the pandemic, widely dispersed WFH productivity surprises drive a lasting shift to WFH. This statement holds even when pre-pandemic expectations about WFH productivity are correct on average.

Second, figure 7 says that pre-pandemic WFH expectations were overly negative for most workers before the pandemic. That is, pandemic-induced experimentation caused most workers to upwardly revise their self-assessed WFH productivity. Online appendix figure A.4 shows that the conditional mean WFH productivity surprise is positive in all twenty-seven countries—ranging up to 8 percent or more in Brazil, India, Italy, Spain, Sweden, Turkey, and the United States. Supposing again that employer and worker assessments are aligned, these revisions in average perceived WFH productivity drive a re-optimization of working arrangements in jobs and tasks on the margin, contributing to a lasting increase in WFH levels. Unlike the "dispersion-of-surprises" effect described in the preceding paragraph, this "average-surprise" effect does not rest on low WFH levels before the pandemic.[19]

To assess whether WFH productivity surprises actually affect WFH levels, we also put the following question to G-SWA participants: "*After COVID, in 2022 and later,* how often *is your employer planning* for you to work full days at home?" The response options are:

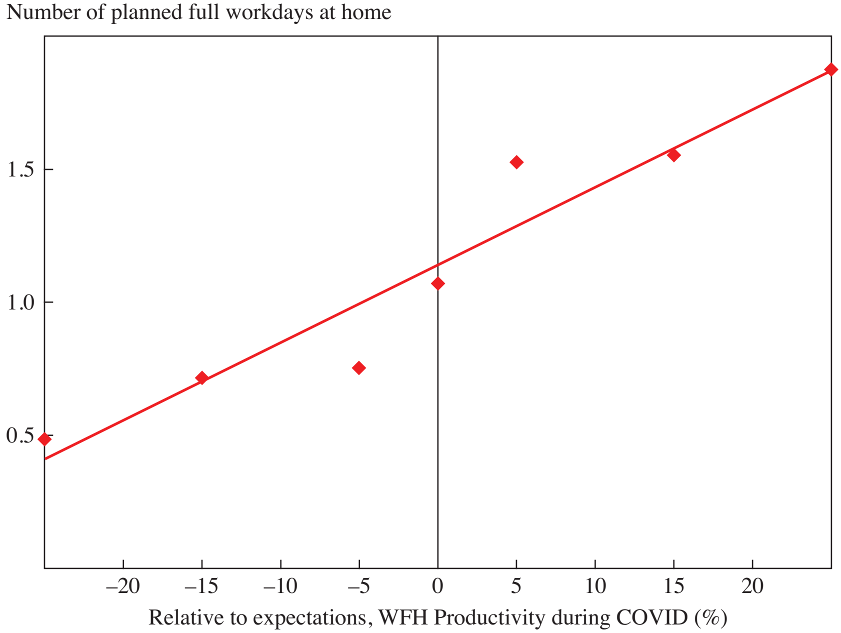
a) Never
b) About once or twice per month
c) 1 day per week
d) 2 days per week
e) 3 days per week
f) 4 days per week
g) 5+ days per week
h) My employer has not discussed this matter with me or announced a policy about it
i) I have no employer

We code response options a, b, and h as zero days and options c through g as one to five days, respectively; we drop persons with no employer from the following analysis.

Figure 8 shows the cross-sectional relationship between employer plans and productivity surprises in the pooled G-SWA data. Planned levels of

19. Because we fielded our surveys 16–23 months after the pandemic's onset, one might worry that worker perceptions of how WFH productivity relates to pre-pandemic expectations are distorted by some form of recall bias. In this regard, we note that Barrero, Bloom, and Davis (2021c) obtain very similar findings in US data for the period from July 2020 to March 2021, much closer to the onset of the pandemic.

Figure 8. Planned Levels of Working from Home after the Pandemic Increase with WFH Productivity Surprises during the Pandemic



Source: Authors' calculations using G-SWA data.
Note: The survey questions read: "Compared to your expectations *before COVID (in 2019)* how has working from home turned out for you?" and "*After COVID, in 2022 and later*, how often *is your employer planning* for you to work full days at home?" The sample consists of 19,027 G-SWA respondents in early 2021 and mid-2022 who worked mainly from home at some point during the COVID-19 pandemic.

WFH *after* the pandemic strongly increase with WFH productivity surprises *during* the pandemic.[20] Moving from the bottom to the top of the surprise distribution involves an increase of about 1.3 days per week in the planned WFH level. Online appendix figure A.5 shows that this strong positive relationship between WFH productivity surprises and planned WFH levels holds in all twenty-seven countries. Barrero, Bloom, and Davis (2021c) find the same strong relationship between WFH productivity surprises and WFH plans using US survey data from July 2020 to March 2021.

20. If primacy bias influences our survey responses, the effect is to attenuate the relationships depicted in figure 8 here, figure A.5 in the online appendix, and the corresponding figure in Barrero, Bloom, and Davis (2021c). This observation follows from the response orderings in the questions that elicit the data behind these figures.

The evidence in figures 7 and 8, and online appendix figure A.5, provides powerful support for our three-part explanation of how and why the pandemic catalyzed a large, lasting uptake in WFH. First, the pandemic drove a mass, compulsory experiment in WFH. Second, mass experimentation generated new information and shifted perceptions about the feasibility and productivity of WFH. Third, the shift in perceptions caused a re-optimization of working arrangements, which included a large, lasting shift to much higher WFH levels. The preconditions for the shift were also in place: major advances during previous decades in the technologies, infrastructure, and products that support the internet, two-way video, and other forms of remote interaction.

This explanation and the supporting evidence do not imply that the big shift to WFH raised productivity. To see this point, consider a simple example of how a shift in perceptions alters the extent of WFH and productivity. Before the pandemic, suppose all workers and their employers perceive WFH to be 10 percent less productive than on-site work. Suppose, as well, that all workers are willing to accept a 5 percent pay discount to WFH. No one works from home in these circumstances, because the perceived productivity loss exceeds the willingness to pay. Now consider what happens in reaction to a pandemic that forces employers and workers to WFH for weeks or months. Based on their experiences during the pandemic, suppose half of workers (and their employers) learn that WFH is about as (un)productive as expected, while the other half learns that it is Δ percent more productive than expected.

There are three interesting cases: (1) When $0 < \Delta < 5$, WFH levels return to zero after the pandemic ends. In this case, the positive productivity surprise is too small to trigger a lasting change in working arrangements. (2) When $5 < \Delta < 10$, half of workers stick with WFH after the pandemic ends, because they now face a productivity discount of only $10 - \Delta$ percent, which is smaller than their willingness to pay to WFH. In this case, the productivity surprise triggers a lasting shift to WFH and a productivity fall of $0.5\ (10 - \Delta)$ percent. For example, if the pandemic leads half of workers to conclude that WFH is only 2 percent less productive than on-site work ($\Delta = 8$), then economy-wide productivity falls 1 percent. (3) When $\Delta > 10$, the productivity surprise drives a lasting shift to WFH and a productivity rise of $0.5\ (\Delta - 10)$ percent. Thus, when forced experimentation leads to a lasting shift to WFH, it can bring higher or lower productivity.

Drawing on additional data for the United States, Barrero, Bloom, and Davis (2021c) estimate that the lasting shift to WFH raised the economy-wide level of labor productivity by about 1 percent. The productivity effect could be larger or smaller in other countries, and it could well be negative in some countries. Indeed, it could be negative in some industries and regions within the United States, even if it's positive on average.

Our three-part explanation for the big shift also addresses another question: If WFH is now attractive for many employees and organizations, why did the shift not happen sooner and more gradually? Our answer is that the full benefits of WFH went unrecognized and unrealized before the pandemic drove a sudden, huge surge in experimentation that led to major revisions in perceptions about the feasibility and productivity of WFH. The simultaneity of large-scale experimentation is important in this regard. A law firm, for example, could have experimented with WFH before the pandemic. What it could not have done was experiment with WFH when the courts and other firms—including clients, rival law firms, consultants, and suppliers—also worked remotely. Had the COVID-19 pandemic not occurred, our evidence suggests that the big shift to WFH would have taken place much more slowly over many years.

Emanuel and Harrington (2021) highlight a different explanation for why remote work was rare before the pandemic: employers were reluctant to offer remote-work jobs because those jobs attracted less able employees. To assess the empirical relevance of this selection effect, Emanuel and Harrington (2021) study call center employees of a major online retailer. They find that more able people tend to favor on-site work to improve their promotion prospects and to avoid pooling with less productive coworkers. This type of selection effect in the relationship between worker ability and work mode (remote or on-site) can deter employers from offering remote work, even when remote work does not hurt productivity for any given worker. As Emanuel and Harrington (2021) recognize, this explanation for the rarity of remote work before the pandemic does not explain the pandemic's role in catalyzing a lasting uptake in WFH. In the context of their sorting model, explaining the lasting uptake in WFH also requires an improvement in the capacity of employers to screen workers or an increase in preference heterogeneity over WFH.

III.B. Other Forces That Helped Propel a Lasting Shift to WFH

Several other forces helped propel a lasting shift to WFH. One such force is the change in social attitudes regarding WFH. To investigate this matter, we asked G-SWA respondents the following: "Since the COVID pandemic began, *how have perceptions about working from home (WFH) changed among people you know?*" The response options are:

a) Hugely improved—the perception of WFH has improved among almost all (90–100%) the people I know (95%)

b) Substantially improved—the perception of WFH has improved among most, but not all, of the people I know (70%)
c) Slightly improved—the perception of WFH has improved among some people I know but not most (25%)
d) No change (0%)
e) Slightly worsened—the perception of WFH has worsened among some, but not most, people I know (–25%)
f) Substantially worsened—the perception of WFH has worsened among most, but not all, people I know (–70%)
g) Hugely worsened—the perception of WFH has worsened among almost all (90–100%) the people I know (–95%)

We use the percentage values in parentheses to assign a numerical score to each response; these percentage values did not appear in the questionnaire.

Applying the same regression approach as before to these numerical scores, online appendix figure A.6 reports evidence that the social acceptance of WFH has risen sharply in all countries since the pandemic.[21] Thus, those who WFH are much less likely to be seen as shirkers and slackers now than before the pandemic. As a result, managers have become more willing to offer WFH to retain and recruit employees.[22] Employees who value WFH are now less hesitant to work remotely when given the chance. In this way, the dramatic improvement in the social acceptance of WFH contributes to the size and stickiness of the big shift to WFH.

Several studies provide evidence of other forces that helped drive and entrench the big shift to WFH. Riom and Valero (2021) and Eberly, Haskell, and Mizen (2021) present evidence that the pandemic prompted firms to invest in new workplace equipment and new digital technologies that support remote work. Barrero, Bloom, and Davis (2021c) use SWAA data to quantify capital investments at home in response to the pandemic and worker time devoted to learning how to WFH. They estimate the value of these pandemic-induced investments at 0.7 percent of annual GDP. Criscuolo and others (2021) and Riom and Valero (2021) present evidence

21. Barrero, Bloom, and Davis (2021c) find the same strong result for the United States in SWAA data. Moreover, the result has persisted for more than two years since the onset of the pandemic in repeated cross sections of SWAA data. See the updates at WFH Research, "Working from Home before and since the Start of COVID," http://www.wfhresearch.com/. Thus, there's little reason to think that the increase in the social acceptance of WFH will reverse anytime soon, if ever.

22. Davis, Macaluso, and Waddell (2022) provide direct evidence that many employers now offer remote work to retain and recruit employees based on a survey conducted by the Federal Reserve Bank of Richmond in late 2021.

that firms adopted new managerial practices to support WFH in reaction to the pandemic. Bloom, Davis, and Zhestkova (2021) find that, in the wake of the pandemic, new patent applications shifted toward technologies that support WFH and remote interactions more generally. All of these various investments in equipment, skills, technologies, and managerial practices create durable forms of capital and knowledge that improve performance in the WFH mode now and in the future. In addition, Barrero, Bloom, and Davis (2022a) present SWAA-based evidence that the pandemic created long-lingering concerns about infection risks among some workers and that these concerns, in turn, led some workers to prefer jobs that allow WFH.

There is another force—a strategic complementarity—that amplifies the direct effects of all the other forces discussed above, including the effects of experimentation, learning, and re-optimization. Specifically, WFH becomes more attractive relative to work in the office when a larger share of coworkers also works remotely. This force operates most clearly in the extreme: when no one else works in the office, there's no point in commuting to reap the benefits of face-to-face interactions. This type of strategic complementarity also operates at the level of organizations. As an example, it makes more sense for a law firm to allow or encourage partners, associates, and other staff to WFH when clients also work remotely. In short, WFH makes more sense when others WFH than when everyone works on business premises.

IV. Societal Experiences and Post-pandemic WFH Levels

We now investigate how societal experiences during the pandemic have affected employer plans regarding WFH in the post-pandemic economy and other outcomes. We consider two aspects of societal experience. First, the cumulative stringency of government-mandated restrictions on commercial and social activities during the pandemic, or cumulative lockdown stringency as a shorthand. Second, the severity of the pandemic itself, as measured by cumulative COVID-19 death rates.

To measure lockdown stringency, *LS*, we draw on the widely used Oxford data described in Hale and others (2021).[23] For each country (or region within a country), we construct an index that combines the extent and duration of government restrictions on commercial and social activity,

23. The data are available at University of Oxford, Blavatnik School of Government, "COVID-19 Government Response Tracker," www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker.

following the approach in Baker, Davis, and Levy (2022). As a first step, we compute the monthly lockdown stringency value for country c in month t as:

$$(1) \qquad LS_{ct} = \text{Max}\left\{\text{SIPO}, \left(\frac{3}{4}\right)\text{BCO} + \left(\frac{1}{4}\right)\text{SCO}\right\},$$

where SIPO = 1 when a shelter-in-place order is in effect, zero otherwise; BCO = 1 when a broad-based business closure order is in effect; and SCO = 1 when schools are closed. These indicator variables take fractional values when the order is in effect part of the month or in part of the country. In a second step, we cumulate the *LS* values from March 2020 through the month before the survey wave for the country in question to obtain our cumulative lockdown stringency (CLS) index. This index summarizes the extent and duration of government restrictions on economic and social activity through the month before the survey wave.

We measure cumulative COVID-19 deaths per capita through the end of the month before the survey wave. Our data on reported COVID-19 deaths are from the Johns Hopkins Coronavirus Resource Center.[24] Some argue that excess mortality measures are more appropriate for many purposes than reported COVID-19 deaths. There is merit in this argument. However, excess mortality measures of COVID-19 fatalities are unavailable for some countries, and they can be sensitive to the statistical procedure used to define the excess concept. In light of these facts, we use reported deaths from an authoritative source.

Online appendix figures A.7 and A.8 show the country-level values of our CLS index and cumulative COVID-19 death rates per capita. There is a great deal of cross-country variation in these measures, which is useful in our efforts to assess how cumulative lockdown stringency and cumulative COVID-19 deaths relate to planned WFH levels and other outcomes.

To assess whether pandemic severity and lockdown stringency help explain country-level differences, we fit unweighted least squares regressions of the following form to individual-level G-SWA outcomes:

$$(2) \qquad Y_{icw} = \gamma^{PS} PS_{icw} + \gamma^{LS} CLS_{icw} + X_{icw}\beta + \varepsilon_{icw},$$

where PS_{icw} and CLS_{icw} are the cumulative pandemic severity and lockdown stringency measures, respectively, for person i in country c and survey wave w. The X_{icw} vector contains our individual-level controls for gender, four age groups, three education groups, and eighteen industry sectors plus wave fixed effects and the national value of log real GDP per capita.

24. Johns Hopkins University Coronavirus Resource Center, https://coronavirus.jhu.edu.

Table 2. Current and Planned Levels of Working from Home Rise with the Cumulative Stringency of Government-Mandated Lockdowns

	Outcome			
	Current WFH days per week	*Desired WFH days per week*	*Planned WFH days per week*	*Amenity value of option to WFH two to three days per week*
	(1)	*(2)*	*(3)*	*(4)*
Cumulative lockdown stringency	0.204**	0.085	0.136***	0.363
	(0.078)	(0.057)	(0.047)	(0.418)
Cumulative COVID-19 deaths per capita	–0.005	0.044	–0.039	0.263
	(0.086)	(0.059)	(0.056)	(0.299)
Observations	33,091	36,078	34,875	36,078
R^2	0.098	0.069	0.086	0.057

Source: Authors' calculations using G-SWA data.

Note: All regressions include controls for log real GDP per capita, gender, four age groups, three education groups, eighteen industry sectors, and wave fixed effects. The reported COVID-19 deaths and lockdown stringency measures are standardized to zero mean and unit standard deviation across countries. Errors clustered at the country level.

$^{**}p < .05$, $^{***}p < .01$.

Table 2 reports our first set of regression results. Greater levels of the CLS index are associated with positive and statistically significant effects on current WFH levels (as of the survey) and post-pandemic planned levels of WFH.[25] Column 3 implies that an increase in the CLS index value equal to two standard deviations (across countries) raises the number of planned WFH days by 0.27 days per week. That amounts to about 38 percent of the cross-country mean WFH plan reported in figure 2. We find no statistically significant effect of CLS on desired WFH levels or on the WFH amenity value. We find no statistically significant effect of cumulative COVID-19 death rates on *any* of the outcome variables in table 2.

Expanding the specifications to include a measure of cumulative mask mandates has no impact on the estimated CLS effect on planned WFH days, as reported in online appendix table A.5. Whether mask mandates should be seen as a milder form of social restrictions or as conceptually different from the other restrictions covered by our CLS index is unclear.

25. Criscuolo and others (2021) find that firms in countries with stricter lockdown measures in spring 2020 had higher WFH levels at the time conditional on sector and firm-size fixed effects and each firm's pre-pandemic WFH level; see their table A.3 and related discussion. This result is consistent with our results but quite distinct. Whereas they find that WFH levels in the early stages of the pandemic rose with contemporaneous lockdown stringency, we find that future WFH levels rise with cumulative lockdown stringency during the pandemic in surveys conducted 16–23 months after the pandemic's onset.

The table also provides evidence that mask mandates, unlike lockdowns, raise desired WFH days and the amenity value of the option to WFH. These results are consistent with the two-part idea that first, (many) people dislike wearing masks on the job, and second, compelling them to do so leaves a residue of distaste for working on business premises.

Adapting the specifications to encompass regional variation where available yields somewhat larger effects of the CLS index on current WFH days and somewhat smaller effects on planned WFH days (online appendix table A.6). We also tried replacing our CLS index with a cumulative version of the index in Hale and others (2021). Relative to our index, theirs uses additional inputs that pertain to the cancellation of public events, restrictions on gathering size, public transport closures, restrictions on internal movements and international travel, and public information campaigns. These additional inputs are hard to measure in some countries, and public information campaigns are conceptually distinct from activity restrictions. So there are trade-offs between using our CLS index and our cumulative version of their broader index. As it turns out, results are very similar when using their index in place of ours.

Finally, we rerun the regression specifications in table 2 on samples limited to (a) all college-educated persons and (b) all persons with a postgraduate degree. As reported in table 3, the estimated lockdown effects on current and planned WFH levels are larger when we limit the sample to college-educated persons. They are larger yet when we focus on graduate-degree holders. Specifically, relative to the full-sample results in table 2, the estimated effects of the CLS index on current and planned WFH levels are twice as large for graduate-degree holders. In unreported results, we find the same pattern in limited-sample analogs to online appendix tables A.5, A.6, and A.7. Greater sensitivity to lockdown stringency among workers with more education is perhaps no surprise because they are more likely to hold jobs for which remote work is feasible.

To summarize, employers plan higher post-pandemic WFH levels in countries and regions with greater cumulative restrictions on commercial and social activities during the pandemic, conditional on a battery of controls.[26] This result suggests that employers more fully adapted their business models and personnel practices to remote work in countries that imposed more stringent lockdowns. Such a response could arise via learning-by-doing effects, whereby more experience with strict lockdowns leads to fuller

26. Evidence on daily stock market reactions to government lockdown announcements supports the view that the lockdowns themselves had material effects on economic activity; see Ashraf (2020) and Yang and Deng (2021).

Table 3. Lockdown Effects Are Stronger for the More Educated

	Outcome			
	Current WFH days per week	*Desired WFH days per week*	*Planned WFH days per week*	*Amenity value of option to WFH two to three days per week*
	(1)	*(2)*	*(3)*	*(4)*
A. Restricting the sample to persons with a college degree				
Cumulative lockdown stringency	0.282***	0.092	0.170**	0.503
	(0.097)	(0.067)	(0.064)	(0.433)
Cumulative COVID-19 deaths per capita	–0.037	0.035	–0.059	0.337
	(0.106)	(0.075)	(0.066)	(0.347)
Observations	22,210	24,054	23,317	24,054
R^2	0.085	0.058	0.075	0.049
B. Restricting the sample to persons with a graduate degree				
Cumulative lockdown stringency	0.410***	0.144**	0.266***	0.380
	(0.139)	(0.059)	(0.086)	(0.401)
Cumulative COVID-19 deaths per capita	–0.113	–0.025	–0.105	0.180
	(0.118)	(0.055)	(0.075)	(0.335)
Observations	10,954	11,826	11,468	11,826
R^2	0.082	0.056	0.088	0.036

Source: Authors' calculations using G-SWA data.

Note: This table uses the same specifications and measures as table 2. Errors clustered at the country level.
$^{**}p < .05$, $^{***}p < .01$.

adaptation. It could also arise as a proactive response by employers that see a history of lockdown stringency as predictive of more stringent lockdowns during future infectious disease outbreaks. Another possible interpretation is that more fearful reactions to the pandemic drove more voluntary adoption of remote work practices in some countries *and* more stringent lockdown policies. Here as well, learning-by-doing effects would lead naturally to higher future WFH levels in the more fearful countries that accumulated more WFH experience during the pandemic.

In contrast to the lasting effects of lockdown stringency on current and future WFH levels, we find no evidence that cumulative COVID-19 death rates affect employer plans for post-pandemic WFH levels or current WFH levels as of the survey date.[27] We are surprised by this result, but it appears

27. WFH levels covary positively with the incidence of COVID-19 across US states in April and May 2020 (Brynjolfsson and others 2020), but this pattern is not at odds with our evidence, since it pertains to the relationship of WFH levels to contemporaneous COVID-19 death rates rather than the long-term effects of cumulative COVID-19 deaths.

to be a robust feature of our data. It also points to a puzzle for the fear-based interpretation of our findings with respect to lockdown stringency: if fearfulness drives country-level differences in lockdown stringency, why do cumulative COVID-19 deaths per capita have no explanatory power for current (as of the survey) and planned WFH levels? The answer, if there is one, must involve some manifestation of fearfulness that is uncorrelated with COVID-19 deaths per capita but, nevertheless, highly correlated with lockdown stringency.

V. Some Implications

V.A. Direct Consequences for Workers and Organizations

Section III presents and reviews several pieces of evidence that people like to WFH. This evidence suggests that the big shift to WFH yields large benefits, on average, for workers and their families. Barrero, Bloom, and Davis (2021c) estimate that planned WFH levels in the US economy deliver aggregate time savings equal to 2 percent of pre-pandemic work hours on an earnings-weighted basis.[28] They find even larger gains in worker welfare using individual-level data on commute times, pre-pandemic WFH days, employer plans for post-pandemic WFH days, and willingness to pay to WFH. Their results do not say that all workers benefit from the shift to WFH, only that the direct effects are large and positive on average. Individuals who highly value daily in-person encounters with work colleagues and those who lose valuable learning and networking opportunities may be worse off. The shift to WFH also has direct effects on the level of productivity, and it can affect the well-being of workers and their families through equilibrium effects on wages and prices, the pace of innovation, and the quality of local public goods.

Section III also presents evidence that preferences around WFH vary greatly across individuals and demographic groups. Regulations that raise WFH costs or restrict the set of WFH options limit the capacity of markets to satisfy these preferences. In this regard, Lockton Global Compliance summarizes new, permanent teleworking regulations since March 2020 in seventeen countries; many of the new regulations raise the costs of remote

28. The 2 percent time savings figure is from Davis (2022) and reflects savings in both commuting time and grooming time. The next draft of Barrero, Bloom, and Davis (2021c) will also account for both.

work, making it less viable.[29] Other new regulations push employers to satisfy employee desires to WFH.[30] That approach raises the societal costs of WFH by forcing it onto employers, even when remote work is poorly suited for their businesses. Especially in economies with fluid labor markets, it is more efficient to accommodate WFH preference heterogeneity via the sorting of workers to employers.

Pre-pandemic laws and regulations also matter. In the European context, for example, visa policies can facilitate or constrict remote work across national borders. In the US context, an employee who works remotely from another state can subject the employer to new state-level payroll taxes, trigger legal obligations to collect taxes on sales into the state, and subject the employer to business income taxes in the state (Jacobs and others 2022). These tax consequences and attendant compliance burdens make it costlier to let employees work from other states, especially when the employer does not already operate there.

For employers, WFH preference heterogeneity presents major strategic choices in personnel management and operations. One possibility is to accommodate preference heterogeneity to maximize the available talent pool, reduce employee turnover, and moderate out-of-pocket compensation costs. As of April/May 2022, about 40 percent of firms in the Survey of Business Uncertainty allow WFH one or more days per week "to keep employees happy and to moderate wage-growth pressures" (Barrero and others 2022, figure 1). Roughly half of American firms in another recent survey offer remote or hybrid working arrangements to help recruit new employees and retain current ones (Davis, Macaluso, and Waddell 2022). Downsides of accommodation include fewer in-person communications, greater operational complexity, and greater challenges in onboarding new employees, mentoring, and sustaining company culture.

29. Lockton Global Compliance, "New Remote Working Legislation around the World [Updated]," June 1; https://globalnews.lockton.com/new-remote-working-legislation-around-the-world/. To pick an example not covered by Lockton, the Ministry of Labor and Social Welfare in Mexico recently issued a draft amendment to its Federal Labor Law that would require employers to ensure and verify that the remote site has "reliable electricity, lighting, ventilation, and ergonomic conditions," provides "a safe workplace that allows for an employee's development and continuity," and meets other conditions; see Palma, Villanueva, and Díaz (2022).

30. Perhaps the most prominent example is legislation that would make WFH a legal right in the Netherlands. The legislation, recently passed by the lower house of the Dutch parliament, would force employers to consider employee requests to WFH and to explain why if the request is denied; see Papachristou (2022).

Another strategic option involves a hang-tough approach that compels most or all employees to work on-site on (almost) all workdays. Elon Musk famously demanded that all Tesla employees work in the office at least forty hours a week or "pretend to work somewhere else." Musk sees particular value in the visible, physical presence of senior employees and questions whether companies with flexible working arrangements can develop new products (Nicholas and Hull 2022; Boyle 2022). The hang-tough approach retains a high intensity of in-person communications and can have important operational advantages, but it also narrows the talent pool, requires a larger physical footprint, raises out-of-pocket compensation costs, and lowers retention rates.

CEO Jeremy Stoppelman makes the case for a fully remote workforce: "At Yelp we made the decision to go remote-first in mid 2020. A big part of our calculus was that employees would strongly prefer cutting their commutes"; "How's it going? Quite well! Internal surveys show high satisfaction and continued productivity from our sales, product and engineering teams. We've hired two remote C-level executives both in geographies with no offices and we've got great access to a diverse talent pool"; "So why does hybrid suck? It forces employees to live near an office (high cost areas) and doesn't get rid of the commute. Also hiring is constrained by geography and you have to maintain underutilized office space."[31]

As the foregoing remarks indicate, the trade-offs associated with these three broad strategies—accommodation, hang tough, and fully remote—differ across organizations and workforces and, of course, across industries and occupations. Put another way, there is much heterogeneity on the labor demand side in the capacity to efficiently supply the WFH options that many employees value. Given this demand-side heterogeneity and the supply-side heterogeneity in preferences, a market-based approach to the determination of working arrangements is likely to yield much diversity in WFH outcomes—including many people who never WFH, some who WFH much of the time, others who WFH almost all the time, and employers that adopt a range of accommodation, hang-tough, and fully remote personnel practices. This type of market diversity satisfies heterogeneous WFH preferences in a cost-effective manner. It also lets employers and workers adjust over time in response to their own experiences, learning

31. Twitter, Jeremy Stoppelman, May 24, 2022, https://twitter.com/jeremys/status/1529164087547944960.

from others, and new conditions. Prescriptive regulatory approaches are unlikely to satisfy a broad range of WFH preferences in an equally cost-effective manner.

V.B. WFH and the Pace of Innovation

Historically, many forms of invention, innovation, and entrepreneurship were highly concentrated in space.[32] This empirical regularity gives rise to concerns that the big shift to WFH will slow the pace of innovation. On this front, we see good reasons for optimism. As a first observation, many of the most productive and innovative firms in the world operate across multiple locations, cities, and countries. So, workforce dispersal per se is an unlikely killer of innovation and productivity growth. Stronger grounds for concern rest, instead, on the potential loss of the innovation benefits that flow from gathering a critical mass of creative people in a single location or set of locations in close physical proximity.

Second, key developments that facilitated the big shift to WFH—for example, the rise of the internet, better broadband infrastructure, improved video technologies, and the emergence of the cloud—create greater reach and higher quality in one-way and two-way communications at a distance. In this regard, Pearce (2020, fig. 3) shows that the geographic dispersal of collaborative innovations, as measured by the locations of named inventors in US patent filings, has been rising for decades. Chen, Frey, and Presidente (2022) use author locations to document a similar pattern in scientific publications. They also study the relationship of remote collaboration to the quality of scientific articles, as reflected in citations. Before 2010, remote collaboration produced articles that were more incremental and less likely to yield "disruptive" advances. This quality discount on remote-collaboration articles shrinks over time, vanishes around 2010, and then becomes a premium. A plausible explanation is that advances in remote-collaboration technologies have made it easier and cheaper to coordinate a broader range of specialized and geographically scattered complementary inputs. In the model of Becker and Murphy (1992, sect. 6) such a fall in coordination costs raises the innovation rate.

Yang and others (2022) investigate how the pandemic-induced shift to remote work altered communications among 61,182 Microsoft employees from December 2019 to June 2020. They find that communications became

32. See Carlino and Kerr (2015) and Combes and Gobillon (2015) for reviews of the extensive literature on this topic.

more asynchronous after the shift to remote work and collaborations became more static and siloed. These types of changes can impede the diffusion of knowledge within an organization and slow the pace of innovation. However, the larger implications of their study are unclear for two reasons: organizations that stick with remote work will adapt their practices over time to mitigate the disadvantages and exploit the advantages, and as the pandemic recedes, organizations have strong incentives to revert to in-person collaboration in situations where remote work is ineffective. For both reasons, the near-term impact of a surprise, compelled, and pervasive shift to remote work is a doubtful guide to the longer-term innovation effects of voluntary remote-work adoption.

Third, the big shift to WFH stimulates advances in technologies that facilitate productive interactions at a distance, as suggested by the analysis of new patent applications in Bloom, Davis and Zhestkova (2021). Fourth, and related, the rise of remote work and professional interactions at a distance during the pandemic have overturned customs and practices that, before the pandemic, impeded the flow of ideas and prevented a fuller realization of agglomeration benefits. To take an example that *BPEA* conference participants will readily appreciate, many scientific and professional conferences that once operated in a closed, in-person, invitation-only manner are now partly or fully open to virtual participants. While fewer (or different) people may choose to participate in person, and virtual participation may be less rewarding, opening the door to virtual participation can greatly expand the reach of participation and accelerate the diffusion of ideas.

Fifth, business and managerial practices will adapt to a world of remote work and better technologies for communication at a distance. Tu and Li (2021) offer practical ideas for how organizations can foster mentorship and professional networking and improve rapport between managers and employees in a virtual work setting. Larson, Vroman, and Makarius (2020) stress the need for clear “rules of engagement” in remote work to set ground rules and manage employee expectations. Both articles highlight the need to consciously facilitate social interactions among employees, which surely warrants greater managerial attention in a hybrid or fully remote work environment than in the traditional on-site environment.

We summarize as follows: the scope for positive agglomeration spillovers in virtual space is expanding, even as the shift to WFH diminishes agglomeration spillovers in physical space. A full picture of how these countervailing forces affect the pace of innovation is not yet available, but there are good reasons for optimism.

V.C. Challenges for Cities

There are stronger reasons for concern when it comes to the fortunes of cities.[33] The big shift to WFH presents especially acute challenges for dense urban centers that are organized to support a large volume of inward commuters and a high spatial concentration of commercial activity. Consider a few statistics that speak to the scale of the challenge: WFH accounts for 38 percent of full paid workdays in the ten most populous US metro areas as of June 2022, as compared to 30 percent in the next forty most populous areas, and 27 percent in smaller cities and towns (Barrero, Bloom, and Davis 2022b, slide 15). The share is nearly 45 percent in the San Francisco Bay area. These WFH levels are at least 20–30 percentage points above pre-pandemic levels. They have also stabilized in recent months, which suggests they are here to stay.

Ozimek and O'Brien (2022) document some sobering developments regarding population flows. From 2020 to 2021, population fell in 68 percent of urban counties that intersect an urban area with at least 250,000 people. Children under age 5 in urban counties fell 3.7 percent from 2020 to 2021, as compared to 2.4 percent nationwide. The most populous urban areas saw especially large drops. San Francisco lost 7.6 percent of its under-5 population from 2020 to 2021 and more than 10 percent from 2019 to 2021. In contrast, the under-5 population shrank more slowly from 2010 to 2019 in urban counties than across the nation as a whole. These observations support the view that newfound opportunities to WFH raise the attractiveness of suburban and exurban living, especially for families with young children that seek lower housing costs and better schooling options. Rising murder rates in many US cities (Elinson 2022) are another factor contributing to urban outmigration, again facilitated by the rise of WFH.

Real estate markets tell a consistent story. Rosenthal, Strange, and Urrego (2022) examine 68,000 newly executed commercial leases across eighty-nine US cities from January 2019 to October 2020. They find that the elasticity of rental values with respect to employment density fell 2 percentage points in the wake of the pandemic. Large, dense cities that rely heavily on subway and light rail also saw a 15 percent fall in the commercial rent gradient (distance from city center) and a decline in the transit rent premium. Gupta, Mittal, and Van Nieuwerburgh (2022) combine data on commercial lease revenues, office occupancy rates, and market rents with an

33. We focus here on challenges to cities in rich countries, especially the United States. As Edward Glaeser points out in his comment on this paper, cities in poor countries face a somewhat different set of challenges.

asset-pricing model to estimate that the pandemic-induced shift to remote work drove a 45 percent drop in office values in 2020 and a 39 percent drop in the longer run. Ramani and Bloom (2021) use Zillow home value indexes to examine residential real estate prices. Their figure 1 shows that home values in central business districts fell 2 percent in nominal terms from February 2020 to April 2021, 7 percent relative to prices in the top decile of zip codes by population density, and 13 percent relative to prices in the next four deciles.

One important implication of these developments is that the big shift to WFH drove a large, persistent negative shock to the local tax base in many cities. Fewer inward commuters means a smaller sales tax base, as does residential outmigration. Fewer inward commuters lowers transit revenues. The incomplete recovery of business travel means lower hotel occupancy tax revenues. The fall in real estate values erodes the local property tax base. All of these fiscal effects tend to be more intense in denser urban areas.

Glaeser, Kolko, and Saiz (2001) and Florida (2012) argue that cities become, and remain, successful by offering lifestyle and consumption opportunities that people value. The big shift to WFH makes urban amenities even more important for city success, because the ability to WFH two or three days a week lowers the cost of residing far from a job that, nominally, is located in the city. For those who can WFH four or five days a week, the pressure to live close to work is weaker still. Cities that do not provide good schools, do not control crime, levy high taxes, and do not provide attractive places for people to live, work, and play are now more exposed to residential outmigration and big drops in inward commuting. They now face greater risks of a downward spiral in local tax revenues and urban amenities. (By a similar logic, attracting good jobs will do less to boost urban fortunes when those jobs can be performed elsewhere much of the time.) The flip side of these observations is that cities and suburbs that offer good schools, low crime, and pleasant places to live, work, and play are even more attractive now than before the pandemic.

That brings us to the second important implication for cities: the rise of remote work raises the elasticity of the local tax base with respect to the quality of local governance—more so in cities like San Francisco where so many well-paying jobs are amenable to remote work. This increase in the tax-base elasticity creates sharper incentives for sensible, efficient local governance, which could well yield better management and outcomes in many cities. At the same time, it creates greater scope for a downward spiral in city fortunes, whereby poor governance amplifies outmigration and the loss of inward commuters, eroding the local tax base and undercutting the fiscal capacity to supply local public goods, which then leads to more outmigration and less inward commuting, and so on. In this way, the big shift to WFH has the potential to amplify the negative effects of poor governance, political instability, and crime on the fortunes of cities.

Glaeser (2022) expresses similar concerns, arguing that the COVID-19 pandemic endangers cities because it exacerbates "existing challenges, including adapting to virtual life and the political instability associated with growing urban discontent. . . . The pandemic has also hit cities during a period of discontent over gentrification, racial disparities in policing and inequality more generally, and that creates political risks. . . . If cities try to target their wealthier residents and businesses or if those cities allow urban crime levels to soar, then those taxpayers could easily leave, which in turn could generate a downward spiral, reminiscent of many American cities during the 1970s" (4–5).

Another, related implication: the fallout from the big shift to WFH will differ greatly across cities for multiple reasons. First, the extent of the initial pandemic-induced shift to WFH and hence the size of the negative fiscal shock, differs greatly. Second, property prices and rents will adjust to preserve full use of structures and space in cities with intrinsically strong fundamentals and good governance, even as marginal cities experience a long-term rise in vacancy rates and empty spaces. Third, cities differ in their political capacity to adjust to the WFH shift and the now-greater mobility of well-educated, highly paid workers and the companies that employ them. A larger elasticity of the local tax base with respect to urban amenities and local governance quality may foster better governance in some cities and a downward spiral in others. Fourth, cities that are well endowed with consumer amenities are now in an even better position to attract high-income workers.

The risk that city-level fortunes will diverge is more acute in the United States than in most other rich countries, in part because political decisions about the provision of local public goods are more decentralized in the United States and local fiscal resources are more closely tied to local economic prosperity. These aspects of federalism give rise to more scope for a downward spiral in city-level fiscal resources and urban amenities. Compared to most other countries, the United States also offers more location options with the same language, similar cultures, a similar legal system, and so on. Thus, if governance fails in one city, it is easier to relocate to a better-performing but otherwise similar city. In addition, urban crime levels are higher in the United States than in most other rich countries. Thus, the potential for high or rising crime rates to accelerate a downward spiral in urban fortunes looms larger in the American context.

In short, the big shift to WFH and the now greater sensitivity of local fiscal resources to the quality of local amenities create major challenges for large cities. A failure to meet these challenges would lead to much economic and social harm and at least partly offset the large, direct benefits of WFH discussed above. Moreover, the harms that arise from a failure of (some) cities to adapt to the big shift would be concentrated among poorer households, who have less capacity to move away from urban problems and who also reap smaller direct benefits from the big shift to WFH.

VI. Concluding Remarks

The COVID-19 pandemic catalyzed a large and enduring uptake in work from home bringing major lifestyle changes to millions of workers, a scramble to adapt managerial and personnel practices, major operational challenges for organizations that embrace hybrid or fully remote working arrangements, the redirection of worker spending away from city centers, declines in urban real estate values, and outmigration from some cities. The broader economic and social consequences will unfold for many years to come.

As for how the pandemic catalyzed the big shift to WFH, and why it did not happen sooner and more gradually, we advance a three-part explanation: First, the pandemic compelled a mass social experiment in WFH. Second, that experimentation generated a tremendous flow of new information about WFH and greatly altered perceptions about its practicality and effectiveness. Third, in light of this new information and shift in perceptions, individuals and organizations re-optimized, choosing much more WFH than before the pandemic. We find strong support for this three-part explanation when looking across individuals in the twenty-seven countries covered by our survey. Specifically, the number of full WFH days per week that employers plan after the pandemic rises strongly with employee assessments of WFH productivity surprises during the pandemic. Exploiting cross-country variation, we also find evidence that longer, stricter government lockdowns during the pandemic led to higher WFH levels as of mid-2021 and early 2022 and higher planned WFH levels after the pandemic ends.

Though scattered across many papers (including this one), there is now much evidence that the pandemic also spurred other developments that helped drive a lasting shift to WFH: new investments in the home and inside organizations that facilitate WFH, learning by doing in the WFH mode, advances in products and technologies that support WFH, much greater social acceptance of WFH, and lingering infection concerns that lead some people to prefer remote work. The rise of the internet, emergence of the cloud, and advances in two-way video before the pandemic created the conditions that made possible a big shift to WFH. Thus, the full story of how the pandemic led to a large, lasting shift to remote work has many elements.

We also develop evidence that the shift to WFH benefits workers. The reason is simple: most workers value the opportunity to WFH part of the week, and some value it a lot. It's easy to see why. WFH saves on time and money costs of commuting and grooming, offers greater flexibility in time management, and expands personal freedom. Few people could WFH before the pandemic. Many can do so now. This dramatic expansion in choice benefits millions of workers and their families. Women, people living with children, workers with longer commutes, and highly educated workers tend to put higher values on the opportunity to WFH.

That does not mean everyone benefits. Some people dislike remote work and miss the daily interactions with coworkers. Over time, people who feel that way will gravitate to organizations that stick with pre-pandemic working arrangements. Another concern is that younger workers, in particular, will lose out on valuable mentoring, networking, and on-the-job learning opportunities. We regard this concern as a serious one but have diffuse priors over whether, and how fully, it will materialize. Firms have strong incentives to develop practices that facilitate human capital investments. Individual workers who value those investment opportunities have strong incentives to seek out firms that provide them. If older and richer workers decamp for suburbs, exurbs, and amenity-rich consumer cities, the resulting fall in urban land rents will make it easier for young workers to live in and benefit from the networking opportunities offered by major cities.

Many observers also express concerns about what the rise of remote work means for the pace of innovation. In this regard, we stress that the scope for positive agglomeration spillovers in virtual space is expanding, even as the shift to WFH diminishes agglomeration spillovers in physical space. How these countervailing forces will affect the overall pace of innovation remains to be seen, but we set forth several reasons for optimism.

The implications for cities are more worrisome. The shift to WFH reduces the tax base in dense urban areas and raises the elasticity of the local tax base with respect to the quality of urban amenities and local governance. These developments warrant both hope and apprehension. On the hopeful side, they intensify incentives for cities to offer an attractive mix of taxes and local public goods. Cities that respond with efficient management and sound policies will benefit—more so now than before the pandemic. On the apprehensive side, the economic and social downsides of poor city-level governance are also greater now than before the pandemic. For poorly governed cities, in particular, the larger tax-base elasticity raises the risk of a downward spiral in tax revenues, urban amenities, workers, and residents.

ACKNOWLEDGMENTS A special thanks to Katharine Abraham and Edward Glaeser for many helpful comments. Thanks also to our editor Jim Stock and the *BPEA* conference participants for their remarks. For the financial assistance that made it possible to conduct our Global Survey of Working Arrangements (G-SWA), we gratefully acknowledge the European Bank for Reconstruction and Development, the Asociación Mexicana de Cultura AC, the Becker Friedman Institute at the University of Chicago, the ifo Institute, King's College London, the Smith Richardson Foundation, and the Templeton Foundation.

References

Adams-Prassl, Abi, Teodora Boneva, Marta Golin, and Christopher Rauh. 2020. "Inequality in the Impact of the Coronavirus Shock: Evidence from Real Time Surveys." *Journal of Public Economics* 189:104245.

Adrjan, Pawel, Gabriele Ciminelli, Alexandre Judes, Michael Koelle, Cyrille Schwellnus, and Tara Sinclair. 2021. "Will It Stay or Will It Go? Analysing Developments in Telework during COVID-19 Using Online Job Postings Data." Working Paper 30. Paris: OECD.

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. 2018. "Intergenerational Mobility and Preferences for Redistribution." *American Economic Review* 108, no. 2: 521–54.

Alipour, Jean-Victor, Harald Fadinger, and Jan Schymik. 2021. "My Home Is My Castle—the Benefits of Working from Home during a Pandemic Crisis." *Journal of Public Economics* 196:104373.

Alsan, Marcella, Luca Braghieri, Sarah Eichmeyer, Minjeong Joyce Kim, Stefanie Stantcheva, and David Y. Yang. 2020. "Civil Liberties in Times of Crisis." Working Paper 27972. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27972.

Angelici, Marta, and Paola Profeta. 2020. "Smart-Working: Work Flexibility without Constraints." Working Paper 8165. Munich: CESifo.

Ashraf, Badar Nadeem. 2020. "Economic Impact of Government Interventions during the COVID-19 Pandemic: International Evidence from Financial Markets." *Journal of Behavioral and Experimental Finance* 27:100371.

Bai, John (Jianqiu), Erik Brynjolfsson, Wang Jin, Sebastian Steffen, and Chi Wan. 2021. "Digital Resilience: How Work-from-Home Feasibility Affects Firm Performance." Working Paper 28588. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28588.

Baker, Scott R., Steven J. Davis, and Jeffrey A. Levy. 2022. "State-Level Economic Policy Uncertainty." *Journal of Monetary Economics* 132:81–99.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis. 2020. "COVID-19 Is Also a Reallocation Shock." *Brookings Papers on Economic Activity,* Summer, 329–71.

Barrero, Jose Maria, Nicholas Bloom and Steven J. Davis. 2021a. "Don't Force People to Come Back to the Office Full Time." *Harvard Business Review,* August 24. https://hbr.org/2021/08/dont-force-people-to-come-back-to-the-office-full-time.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis. 2021b. 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.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis. 2021c. "Why Working from Home Will Stick." Working Paper 28731. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28731.

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

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis. 2022b. "SWAA July 2022 Updates." Slides, July 27. https://wfhresearch.com/wp-content/uploads/2022/07/WFHResearch_updates_July2022.pdf.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis. 2022c. "SWAA September 2022 Updates." Slides, September 9. https://wfhresearch.com/wp-content/uploads/2022/09/WFHResearch_updates_September2022.pdf.

Barrero, Jose Maria, Nicholas Bloom, Steven J. Davis, Brent H. Meyer, and Emil Mihaylov. 2022. "The Shift to Remote Work Lessens Wage-Growth Pressures." Working Paper 30197. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30197.

Bartik, Alexander W., Zoe B. Cullen, Edward L. Glaeser, Michael Luca, and Christopher T. Stanton. 2020. "What Jobs Are Being Done at Home during the Covid-19 Crisis? Evidence from Firm-Level Surveys." Working Paper 27422. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27422.

Becker, Gary S., and Kevin M. Murphy. 1992. "The Division of Labor, Coordination Costs, and Knowledge." *Quarterly Journal of Economics* 107, no. 4: 1137–60.

Berniell, Ines, Leonardo Gasparini, Mariana Marchionni, and Mariana Viollaz. 2021. "The Role of Work-from-Home in the Gender Asymmetries of COVID-19: An Analysis for Latin America Based on High-Frequency Surveys." Working Paper. Washington: World Bank Group.

Bick, Alexander, Adam Blandin, and Karel Mertens. 2020. "Work from Home before and after the COVID-19 Outbreak." Working Paper 2017. Federal Reserve Bank of Dallas. https://www.dallasfed.org/-/media/documents/research/papers/2020/wp2017r2.pdf.

Bloom, Nicholas, Steven J. Davis, and Yulia Zhestkova. 2021. "COVID-19 Shifted Patent Applications toward Technologies That Support Working from Home." *American Economic Association Papers and Proceedings*, vol. 111. Pittsburgh, Pa.: AEA Publications.

Bloom, Nicholas, Ruobing Han, and James Liang. 2022. "How Hybrid Working from Home Works Out." Working Paper 30292. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30292.

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying. 2015. "Does Working from Home Work? Evidence from a Chinese Experiment." *Quarterly Journal of Economics* 130, no. 1: 165–218.

Boyle, Matthew. 2022. "Elon Musk's Demand Staffers Stop 'Phoning It In' May Cost Him Talent." *Bloomberg*, June 1. https://www.bloomberg.com/news/articles/2022-06-01/musk-s-return-to-office-mandate-flies-in-the-face-of-new-reality?sref=v1Ta8kqP&leadSource=uverify%20wall.

Brynjolfsson, Erik, John J. Horton, Adam Ozimek, Daniel Rock, Garima Sharma, and Hong-Yi TuYe. 2020. "COVID-19 and Remote Work: An Early Look at US

Data." Working Paper 27344. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27344.

Carlino, Gerald, and William R. Kerr. 2015. "Agglomeration and Innovation." In *Handbook of Regional and Urban Economics, Volume 5A*, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange. Amsterdam: North Holland.

Chen, Chinchih, Carl Benedikt Frey, and Giorgio Presidente. 2022. "Disrupting Science." Working Paper 2022-04. Oxford: Oxford Martin School, University of Oxford. https://www.oxfordmartin.ox.ac.uk/downloads/academic/Disrupting-Science-Upload-2022-4.pdf.

Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The Empirics of Agglomeration Economies." In *Handbook of Regional and Urban Economics, Volume 5A*, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange. Amsterdam: North Holland.

Criscuolo, Chiara, Peter Gal, Timo Leidecker, Francesco Losma, and Giuseppe Nicoletti. 2021. "The Role of Telework for Productivity during and Post-COVID-19: Results from an OECD Survey among Managers and Workers." Working Paper 31. Paris: OECD.

Davis, Steven J. 2022. "The Big Shift to Working from Home." Slides, NBER Macro Annual Session on "The Future of Work," April 1. https://static1.squarespace.com/static/5e2ea3a8097ed30c779bd707/t/62474133a4e30576426c57c3/1648836921090/Big+Shift%2C+NBER%2C+1+April.pdf.

Davis, Steven J., Claudia Macaluso, and Sonya Ravindranath Waddell. 2022. "Are Firms Using Remote Work to Recruit and Retain Workers?" Economic Brief 22-07. Federal Reserve Bank of Richmond. https://www.richmondfed.org/publications/research/economic_brief/2022/eb_22-07.

Dechezleprêtre, Antoine, Adrien Fabre, Tobias Kruse, Bluebery Planterose, Ana Sanchez Chico, and Stefanie Stantcheva. 2022. "Fighting Climate Change: International Attitudes toward Climate Change." Working Paper 1714. Paris: OECD.

Eberly, Janice C., Jonathan Haskel, and Paul Mizen. 2021. "'Potential Capital,' Working from Home, and Economic Resilience." Working Paper 29431. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29431.

Elinson, Zusha. 2022. "Murders in U.S. Cities Were Near Record Highs in 2021." *Wall Street Journal,* January 6. https://www.wsj.com/articles/murders-in-u-s-cities-were-near-record-highs-in-2021-11641499978.

Emanuel, Natalia, and Emma Harrington. 2021. "'Working' Remotely? Selection, Treatment, and the Market Provision of Remote Work." Working Paper. https://scholar.harvard.edu/eharrington/publications/working-remotely-selection-treatment-and-market-provision-remote-work.

Eurofound. 2020. *Living, Working and COVID-19.* Luxembourg: Publications Office of the European Union.

Florida, Richard. 2012. *The Rise of the Creative Class*. New York: Basic Books.

Glaeser, Edward L. 2022. "Urban Resilience." *Urban Studies* 59, no. 1: 3–35.

Glaeser, Edward L., Jed Kolko, and Albert Saiz. 2001. "Consumer City." *Journal of Economic Geography* 1, no. 1: 27–50.

Gupta, Arpit, Vrinda Mittal, and Stijn Van Nieuwerburgh. 2022. "Work from Home and the Office Real Estate Apocalypse." Working Paper 30526. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30526.

Hale, Thomas, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, and others. 2021. "A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker)." *Nature Human Behavior* 5:529–38.

Hansen, Stephen, Peter John Lambert, Nick Bloom, Steven J. Davis, Raffaella Sadun, and Bledi Taska. 2022. "Remote Work across Jobs, Cities, and Countries." Slides. https://fbe.unimelb.edu.au/__data/assets/pdf_file/0020/4182320/remote_work_presentation_2.pdf.

He, Haoran, David Neumark, and Qian Weng. 2021. "Do Workers Value Flexible Jobs? A Field Experiment." *Journal of Labor Economics* 39, no. 3: 709–38.

Jacobs, Kevin M., Emilio Martinez, Emily L. Foster, and Kaneez Masoom. 2022. "Workforce Changes Compel State Tax Nexus Reassessment." *A&M Bulletin*, February 4. https://www.alvarezandmarsal.com/insights/workforce-changes-compel-state-tax-nexus-reassessment.

Kahn, Matthew E. 2022. *Going Remote: How the Flexible Work Economy Can Improve Our Lives and Our Cities*. Oakland: University of California Press.

Ker, Daniel, Pierre Montagnier, and Vincenzo Spiezia. 2021. "Measuring Telework in the COVID-19 Pandemic." OECD Digital Economy Paper 314. Paris: OECD Publishing.

Kidwai, Aman. 2022. "Spotify Allowed Its 6,500 Employees to Work from Anywhere in the World. Its Turnover Rate Dropped." *Fortune*, August 2. https://fortune.com/2022/08/02/spotify-allowed-6500-employees-work-from-anywhere-in-world-turnover-rate-dropped-remote-work/.

Larson, Barbara Z., Susan R. Vroman, and Erin E. Makarius. 2020. "A Guide to Managing Your (Newly) Remote Workers." *Harvard Business Review*, March 18. https://hbr.org/2020/03/a-guide-to-managing-your-newly-remote-workers.

Lewandowski, Piotr, Katarzyna Lipowska, and Mateusz Smoter. 2022. "Working from Home during a Pandemic: A Discrete Choice Experiment in Poland." Discussion Paper 15251. Bonn: IZA Institute of Labor Economics.

Mas, Alexandre, and Amanda Pallais. 2017. "Valuing Alternative Work Arrangements." *American Economic Review* 107, no. 12: 3722–59.

Nicholas, Katrina, and Dana Hull. 2022. "Elon Musk's Ultimatum to Tesla Execs: Return to the Office or Get Out." *Bloomberg*, June 1. https://www.bloomberg.com/news/articles/2022-06-01/musk-s-tesla-ultimatum-return-to-office-or-work-somewhere-else?sref=v1Ta8kqP&leadSource=uverify%20wall.

Ozimek, Adam. 2020. "The Future of Remote Work." Upwork, May 27. https://www.upwork.com/press/releases/the-future-of-remote-work.

Ozimek, Adam, and Connor O'Brien. 2022. "Families with Young Children Led Exodus from Major Cities during COVID." Economic Innovation Group, July 27. https://eig.org/family-exodus/.

Palma, Leslie, Marité Villanueva, and Alma Díaz. 2022. "Mexico Publishes NOM-037, Draft Health and Safety Conditions for Teleworking." *National Law Review* 12, no. 350.

Papachristou, Lucy. 2022. "Netherlands Poised to Make Work-from-Home a Legal Right." *Wall Street Journal*, July 7. https://www.wsj.com/articles/netherlands-poised-to-make-work-from-home-a-legal-right-11657206737.

Pearce, Jeremy. 2020. "Idea Production and Team Structure." Working Paper. https://static1.squarespace.com/static/5cdc60dc5514b40001c86f7a/t/5f57faad3513452da704e739/1599601354053/Pearce_sep20.pdf.

Ramani, Arjun, and Nicholas Bloom. 2021. "The Donut Effect of COVID-19 on Cities." Working Paper 28876. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28876.

Riom, Capucine, and Anna Valero. 2021. "The Business Response to COVID-19: The CEP-CBI Survey on Technology Adoption." Working Paper 009. London: Center for Economic Performance, London School of Economics and Political Science. https://cep.lse.ac.uk/pubs/download/cepcovid-19-009.pdf.

Rosenthal, Stuart S., William C. Strange, and Joaquin A. Urrego. 2022. "JUE Insight: Are City Centers Losing Their Appeal? Commercial Real Estate, Urban Spatial Structure, and COVID-19." *Journal of Urban Economics* 127:103381.

Saad, Lydia, and Ben Wigert. 2021. "Remote Work Persisting and Trending Permanent." Gallup News, October 13. https://news.gallup.com/poll/355907/remote-work-persisting-trending-permanent.aspx.

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

Tu, Marianna, and Michael Li. 2021. "What Great Mentoring Looks Like in a Hybrid Workplace." *Harvard Business Review*, May 12. https://hbr.org/2021/05/what-great-mentorship-looks-like-in-a-hybrid-workplace.

Wiswall, Matthew, and Basit Zafar. 2018. "Preference for the Workplace, Investment in Human Capital, and Gender." *Quarterly Journal of Economics* 133, no. 1: 457–507.

Yang, Haochang, and Peidong Deng. 2021. "The Impact of COVID-19 and Government Intervention on Stock Markets of OECD Countries." *Asian Economics Letters* 1, no. 4. https://a-e-l.scholasticahq.com/article/18646-the-impact-of-covid-19-and-government-intervention-on-stock-markets-of-oecd-countries.

Yang, Longqi, David Holtz, Sonia Jaffe, Siddharth Suri, Shilpi Sinha, Jeffrey Weston, and others. 2022. "The Effects of Remote Work on Collaboration among Information Workers." *Nature Human Behavior* 6:43–54.

Comments and Discussion

COMMENT BY

KATHARINE G. ABRAHAM This paper is one of several written by Jose Barrero, Nicholas Bloom, and Steven Davis, alone or in collaboration with other coauthors, about the post-pandemic growth in work from home (WFH), the factors that have contributed to this growth, and the economic implications of WFH. They were among the first to recognize the potentially transformative effects of the substantial shift to WFH that has occurred since the spring of 2020. My discussion will focus on questions about the paper's findings and conclusions, but I would like first to express my appreciation for the authors' contributions to our understanding of this important phenomenon.

Prior to the pandemic, WFH was relatively uncommon, but it rose rapidly in the spring of 2020. Based on data from the Global Survey of Working Arrangements (G-SWA), the authors estimate that one to two years after the pandemic's onset WFH averaged 1.5 days per week among workers surveyed in the twenty-seven countries where the G-SWA was fielded. Interestingly, the paper reports higher WFH days in countries that initially implemented more severe COVID-19 lockdowns, suggestive of lasting effects attributable tothe experience with WFH that the lockdowns forced on workers and employers. In many cases, the survey results suggest, the experience of WFH has turned out better than participants would have anticipated. The authors argue that WFH can be expected to persist and grow.

My comments will touch on three issues: (1) how confident we should feel about the survey estimates from the G-SWA, (2) what path WFH might have followed absent the pandemic, and (3) whether WFH ultimately will settle at as high a level as the authors appear to believe.

Brookings Papers on Economic Activity, Fall 2022: 331–360

HOW MUCH CONFIDENCE SHOULD WE PLACE IN THE G-SWA ESTIMATES? As described in the paper, the G-SWA was fielded by Respondi, an international survey company, to members of an online panel in each country. The country samples are not probability samples; rather, Respondi makes use of panels assembled using methods described rather opaquely on their website, aiming for samples that, in the authors' words, are "broadly representative by age, gender, income, and regions within countries." Highly educated individuals are overrepresented in the G-SWA, and the paper is clear that the reported WFH estimates for many countries, though not for the United States, apply only to the highly educated population. Even beyond their relatively high education level, however, the people willing to participate in an online panel could differ from others with the same educational attainment in ways that may affect the survey estimates. The authors cite the robust demand for online surveys administered to pre-recruited online panels in marketing and other commercial applications as evidence supporting their use, but it does not follow from the fact that commercial customers see value in data collected from such panels that they are suitable for producing population estimates.

In addition to concerns about the representativeness of the survey sample, I also have concerns about potential measurement biases in the answers to some of the survey questions. The survey questions that ask for straightforward factual information—how many days per week a person works from home or what their employers have said about future plans for WFH—should be relatively easy for respondents to answer. In contrast, the answers to survey questions that require respondents to make judgments about things they haven't previously considered are more likely to be affected by how the questions are presented.

One potential issue with some of the G-SWA questions is what survey methodologists refer to as primacy bias, the tendency of respondents in self-administered surveys to select answers that appear earlier in a list of possible response options (Groves and others 2009). Consider the key G-SWA question about WFH productivity relative to expectations:

> Compared to your expectations *before COVID (in 2019)* how has working from home turned out for you?
>
> a) Hugely better—I am 20%+ more productive than I expected
> b) Substantially better—I am 10% to 20% more productive than I expected
> c) Better—I am 1% to 10% more productive than I expected
> d) About the same
> e) Worse—I am 1% to 10% less productive than I expected
> f) Substantially worse—I am 10% to 20% less productive than I expected
> g) Hugely worse—I am 20%+ less productive than I expected

The response options for this question are ordered so that positive productivity surprises appear first on the list. Primacy bias could make respondents more likely to select those answers. The G-SWA question about how perceptions of WFH have changed has a similar structure, creating a possible bias toward saying perceptions have improved. Whether primacy bias is a problem for these questions could be tested in future survey waves by varying the response option order.

In addition to the order of the response options, the range of choices provided also can affect how respondents' views or behavior are characterized. When asked a question to which they do not have a ready answer, respondents are likely to look at the range of the response options for clues about a reasonable response. To illustrate with an example from the survey methodology literature, Schwarz and others (1985) asked a sample of people how much time they spent watching television each day. Half of the respondents were randomly assigned to a version of the question with six response options and a top category of two and a half hours or more per day; the other half were given a set of six response options with a top category of four and a half hours or more per day. The latter group was more than twice as likely to say they watched more than two and a half hours of television per day (37.5 percent versus 16.2 percent).

In answering the G-SWA question about how their WFH productivity compared to what they had expected, a majority of survey respondents with WFH experience indicated that they had experienced a positive productivity surprise and most of the remainder said their WFH productivity was about the same as they had expected. The response options available to respondents who viewed their productivity as higher than expected were more than 20 percent higher, 10 to 20 percent higher, and 1 to 10 percent higher. By giving respondents different cues about what might constitute a large, medium, or small surprise, a different set of groupings could have generated a very different estimate for the average productivity surprise. A similar comment applies to the answers about how large of a pay cut the respondents who view WFH as a benefit (a majority of all respondents) would be willing to accept on a job where they could work from home two or three days a week. The options provided on the existing questionnaire are less than 5 percent, 5 to 10 percent, 10 to 15 percent, 15 to 25 percent, 25 to 35 percent, and 35 percent or more, choices that could lead respondents who might not otherwise have done so to contemplate very large pay cuts as something they might accept. Experimenting with different response categories for these questions in future survey waves could be helpful for understanding the sensitivity of the findings to the set of response options offered to respondents.

COMPARISONS WITH OTHER ESTIMATES AND EVIDENCE Given these questions about the representativeness of the G-SWA sample and potential measurement error in the responses to some of the survey questions, I would like to know how estimates from probability-based surveys and evidence from well-identified research studies line up with the G-SWA numbers. As an exploratory exercise, I sought to identify relevant information on WFH for the United States that could help with benchmarking the G-SWA results.

One key estimate from the G-SWA is that, at the time respondents were surveyed in mid-2021 and early 2022, WFH in the United States averaged 1.6 days per week. In June 2022, a question about working from home asking, "In the last 7 days, have you *or any of the people in your household* teleworked or worked from home?" (emphasis added) was added to the Household Pulse Survey fielded by the Census Bureau. Because respondents were answering both for themselves and for others in their household, simply tabulating these responses would yield an upward biased estimate of the extent of WFH. In September 2022, a new question about respondents' own WFH experience was added to the Household Pulse Survey. Respondents could say they worked from home one to two days per week, three to four days per week, or five or more days per week. Under the assumption that the days-per-week categories in the survey question correspond to 1.5 days, 3.5 days, and 5 days per week, respectively, the new Household Pulse Survey data imply that during the September 14–26, 2022, period employed respondents worked from home an average of 1.1 days per week.[1] This is somewhat below but of the same rough order of magnitude as the G-SWA estimate.

The question about telework introduced in the Current Population Survey (CPS) at the start of the pandemic has asked, "At any time in the last 4 weeks, did you telework or work at home for pay because of the coronavirus pandemic?" This question has become increasingly problematic. In June 2020, 31.3 percent of employed persons answered yes, but by September 2022, that had fallen to 5.2 percent.[2] More than two years out from the pandemic's onset, many teleworking respondents likely answered

1. US Census Bureau, "Week 49 Household Pulse Survey: September 14–September 26," tables 7a and 7b, https://www.census.gov/data/tables/2022/demo/hhp/hhp49.html.

2. US Bureau of Labor Statistics, "Labor Force Statistics from the Current Population Survey," table 1, https://www.bls.gov/cps/effects-of-the-coronavirus-covid-19-pandemic.htm#table1.

no because they did not view their telework as related specifically to the pandemic. Happily, new CPS questions about teleworking are to be introduced in October 2022. Respondents will be asked, "At any time LAST WEEK did you telework or work at home for pay?" If a person answers yes, they will be asked how many of their work hours were telework or work at home hours. These questions are not quite the same as the G-SWA questions, but the resulting data should provide another useful point of comparison.

Another G-SWA finding is that US workers expected their employers to reduce WFH from current levels; on average, employed G-SWA respondents in the United States were working at home an average of 1.6 days, but they expected their employers to reduce this to an average of 0.8 WFH days after the pandemic. This is qualitatively consistent with the finding from the Bureau of Labor Statistics' Business Response Survey, fielded from July 27 to September 30, 2021, that 60.2 percent of private sector establishments that had increased telework during the pandemic planned to make the increase permanent (US Bureau of Labor Statistics 2022).

At least for the United States, the G-SWA estimates related to adoption of WFH seem broadly in line with other currently available information. Other G-SWA results will be harder to benchmark. These include the findings on the effects of WFH on workers' productivity, the productivity surprise associated with WFH, and the amount of their pay that workers would be willing to give up to work from home two to three days per week. The paper cites a number of research studies that have produced results consistent with some of these findings, but further research on all of this is needed.[3]

Although beyond the scope of what I was able to do, there would be value in a systematic compilation of the available evidence on WFH, not only for the United States but also for other countries. I suspect, however, that such an exercise would reveal the paucity of information from sources other than the G-SWA regarding even the basic facts on the prevalence of WFH. If nothing else, I hope that readers of the paper will be convinced that WFH is a topic to which both national statistical offices and academic researchers should be paying attention.

WOULD THE WORK FROM HOME TRANSITION HAVE OCCURRED WITHOUT THE PANDEMIC? An important piece of the paper's argument is that the pandemic and resulting lockdowns were key drivers of the jump in WFH that occurred in the spring of 2020 and the subsequent persistence in WFH. It seems

3. Studies cited include, for example, Bloom and others (2015) on the productivity effects of working from home and Mas and Pallais (2017) on workers' willingness to pay for working from home.

clear that the experience with WFH these events have forced on employers have accelerated its adoption. I would not necessarily conclude, though, that absent the pandemic we would have been stuck in a low WFH equilibrium.

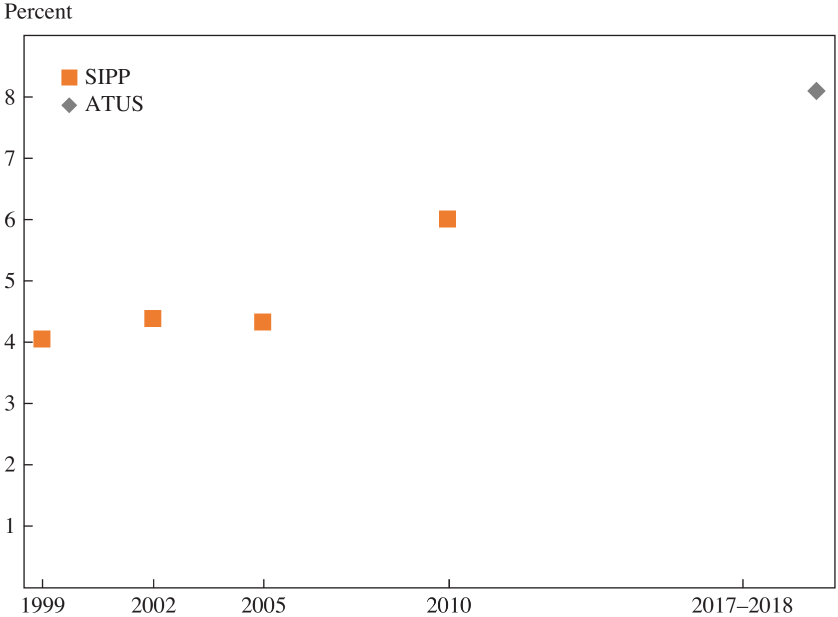
Evidence on the diffusion of technological innovations shows that the spread of new technologies often follows an *S*-shaped pattern, with long periods of very gradual adoption followed by a period of rapid diffusion and then a slowing in growth as adoption approaches its ceiling. Early on, the returns to adoption of a new technology are typically uncertain and diffusion proceeds slowly. At some point, perhaps years or even decades after the introduction of an invention, the value of adoption becomes sufficiently certain that the pace of diffusion accelerates. The classic study of this process is Zvi Griliches's (1957) paper on farmers' adoption of hybrid corn, but there are a multitude of examples in the literature (Hall and Kahn 2003).

The existing data on pre-pandemic WFH unfortunately are sparse but suggest that, at least in the United States, WFH had begun to grow even prior to the pandemic. The Survey of Income and Program Participation (SIPP) collected data on WFH on several occasions between 1995 and 2010, and the American Time Use Survey (ATUS), Leave and Job Flexibilities module, collected data on WFH in 2017–2018. Although one should be cautious about combining data from different sources to track changes over time, the two surveys asked similar questions and both were designed to be nationally representative.[4] Using published data, I was able to construct estimates of the share of wage and salary employees working exclusively from home one day or more a week on their primary job beginning in 1999.[5] As can be seen in figure 1, the SIPP data show this share growing from 4.1 percent in 1999 to 6.0 percent in 2010, with almost all of the

4. In the SIPP, respondents were asked, "As part of the work schedule for [a typical work week during the last month], were there any days when you worked only at home for your job?" If they answered yes, they were asked, "Which days of the week were these?" In the ATUS, Leave and Job Flexibilities module, respondents were asked, "Are there days when you work only at home?" If they answered yes, they were asked, "How often do you work only at home?" The response categories for this second question were five or more days a week, three to four days a week, one to two days a week, at least once a week, once every two weeks, once a month, and less than once a month.

5. The SIPP also included questions about working from home in 1995 and 1997. Because of differences in the way that primary jobs were identified, the 1995 numbers are not comparable to the numbers for later years (Kuenzi and Reschovsky 2001). The estimates published for 1997 did not break wage and salary workers out separately. The overall prevalence of working from home one day a week or more on the primary job as measured in the SIPP was higher than the prevalence for wage and salary workers, but also grew, from 7 percent in both 1997 and 1999 to 9.5 percent in 2010.

Figure 1. Percent of Wage and Salary Employees Working Only at Home at Least One Day per Week, Selected Years, 1999–2018



Sources: Survey of Income and Program Participation; American Time Use Survey, Leave and Job Flexibilities Module.

growth occurring between 2005 and 2010 (Mateyka, Rapino, and Landivar 2012). By 2017–2018, according to the ATUS estimates, the share of wage and salary employees working from home a day or more per week on their primary job had grown to 8.1 percent (US Bureau of Labor Statistics 2019).

Much of the technology that facilitates remote work was developed relatively recently; the World Wide Web, for example, was not invented until 1989, and the software needed for its implementation did not enter the public domain until 1993 (Greenemeier 2009). The timing of the apparent pickup in the pace of growth in WFH suggested by the estimates shown in figure 1 is consistent with the interval between invention and the beginning of widespread adoption for other innovations documented in the literature. To the extent that WFH has real benefits, the history of technological change makes it plausible that, even absent the pandemic, we might have ended up with similar levels of WFH in the not-too-distant future.

HOW SHOULD WE EVALUATE THE PROS AND CONS OF WORKING FROM HOME? The final issue I would like to raise is whether, taking everything into account, the ceiling on the adoption of WFH is as high as the authors suggest. The day-to-day advantages of WFH for workers are highly visible—less time spent getting ready for work, less time spent commuting, and greater flexibility to accomplish other personal and household tasks. What may be less visible are the potential longer-term career costs of WFH. Better technology can help to make remote workers less isolated, but remote work is inherently ill-suited for the informal exchanges of information that are easy when a colleague sits at the next desk or just down the hall. This is likely to be a bigger issue for workers who are new to a firm and especially for those who are just entering the labor market. WFH may prove costly for workers who are unable to develop the professional skills and relationships important to their long-term labor market success. In addition, the appealing flexibility of WFH may be a two-edged sword. The blurring of the distinction between the workday and personal time has undoubted advantages, such as allowing someone to take a few minutes out of their workday to turn over the laundry or meet a contractor at the front door. In the long run, though, WFH may impose costs on workers by leading them to work longer hours and making it more difficult to shut work off outside of normal working hours (Grant, Wallace, and Spurgeon 2013; Felstead and Henseke 2017).

The authors' own work provides one small piece of evidence that some workers may be rethinking their desire for WFH. As noted in the paper, respondents to the separate Survey of Working Arrangements and Attitudes (SWAA) that the authors have been conducting monthly in the United States report a steady rise between January 2021 and June 2022 in the number of WFH days planned by their employers. Over approximately the same period, the SWAA data also show a decline in the number of WFH days workers say they want. Although smaller in magnitude than the increase in WFH days that employers plan, the worker decline accounts for nearly a third of the closing of the gap between employers' plans and workers' desires over that year and a half period.

Like the potential benefits for workers, the potential benefits of WFH for employers also are very visible—being able to recruit from a larger pool of potential workers and, if WFH allows the firm to reduce its physical footprint, saving money on operating expenses. WFH also may increase the productivity of workers performing routine tasks (Bloom and others 2015). My concern is that the loss of informal exchanges of information already mentioned as potentially harmful for workers' careers also may have negative consequences for firms. If WFH impedes collaboration, as I fear is likely to be the case, productivity in the performance of more complex tasks may suffer and innovation may slow.

It is possible that new tools will be developed that can address the challenges to effective remote collaboration. I suspect, however, that remote exchanges will always be an imperfect substitute for in-person interactions. For that reason, I suspect that the ceiling on WFH is not as high as the authors appear to believe. Time will tell which of us is right.

REFERENCES FOR THE ABRAHAM COMMENT

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying. 2015. "Does Working from Home Work? Evidence from a Chinese Experiment." *Quarterly Journal of Economics* 130, no. 1: 165–218.

Felstead, Alan, and Golo Henseke. 2017. "Assessing the Growth of Remote Working and Its Consequences for Effort, Well-Being and Work-Life Balance." *New Technology, Work and Employment* 32, no. 3: 195–212.

Grant, Christine A., Louise M. Wallace, and Peter C. Spurgeon. 2013. "An Exploration of the Psychological Factors Affecting Remote E-Workers' Job Effectiveness, Well-Being and Work-Life Balance." *Employee Relations* 35, no. 5: 527–46.

Greenemeier, Larry. 2009. "Remembering the Day the World Wide Web Was Born." *Scientific American*, March 12. https://www.scientificamerican.com/article/day-the-web-was-born/.

Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* 25, no. 4: 501–22.

Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*, 2nd ed. Hoboken, N.J.: John Wiley and Sons.

Hall, Bronwyn H., and Beethika Kahn. 2003. "Adoption of New Technology." In *New Economy Handbook*, edited by Derek C. Jones. San Diego: Academic Press.

Kuenzi, Jeffrey J., and Clara A. Reschovsky. 2001. *Home-Based Workers in the United States: 1997*. Washington: United States Census Bureau.

Mas, Alexandre, and Amanda Pallais. 2017. "Valuing Alternative Work Arrangements." *American Economic Review* 107, no. 12: 3722–59.

Mateyka, Peter J., Melanie A. Rapino, and Liana Christin Landivar. 2012. *Home-Based Workers in the United States: 2010*. Washington: United States Census Bureau.

Schwarz, Norbert, Hans-J. Hippler, Brigitte Deutsch, and Fritz Strack. 1985. "Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments." *Public Opinion Quarterly* 49, no. 3: 388–95.

US Bureau of Labor Statistics. 2019. "Job Flexibilities and Work Schedules—2017–2018: Data from the American Time Use Survey." News release, September 24. https://www.bls.gov/news.release/pdf/flex2.pdf.

US Bureau of Labor Statistics. 2022. "Over One-Third of Private Sector Establishments Increased Telework during the COVID-19 Pandemic." *TED: The Economics Daily*, October 21. https://www.bls.gov/opub/ted/2022/over-one-third-of-private-sector-establishments-increased-telework-during-the-covid-19-pandemic.htm.

COMMENT BY

EDWARD L. GLAESER In mid-October 2022, workplace visits were down by 22 percent across the United States relative to the start of 2020, according to Google Community Mobility, which produces cell phone–based data. Trips to workplaces had fallen by 39 percent in Manhattan (New York County) and 45 percent in San Francisco. How much time will be spent working from home in the future? Does the persistent emptiness of big city offices today augur a new age in which homes, rather than cubicles, provide the workspaces of the future?

The authors argue that working from home, at least for a few days each week, is likely to become the new normal for large numbers of workers, both in the United States and abroad. They argue that "no other episode in modern history involves such a pronounced and widespread shift in working arrangements in such a compressed time frame." This paper contains a valuable survey that details the spread of working from home in twenty-seven countries. The data in these surveys support the authors' core hypothesis that the relatively low levels of working from home before 2020 reflected a coordination failure. As that failure was remedied by the pandemic, we can look forward to a future where working from home is vastly more common.

I admire both this paper and the authors' broader research enterprise which has tracked working from home since the start of the pandemic, but I am skeptical that this moment in time will ultimately be seen as a watershed that marks the end of the office and factory, especially among less elite workers and especially in the developing world. There is a longer-term trend toward working at home, especially for elite knowledge workers, and that certainly sped up during the pandemic. Moreover, the authors are right to emphasize the "big push" nature of the pandemic event, which unquestionably gave working from home a jolt, and that some of that jolt is permanent.

There are five distinct reasons to be skeptical about any maximalist view of the shift to working from home. First, both electronic surveys and data generated by cell phones are likely to overrepresent technologically connected individuals, and those tech-savvy respondents surely experienced far more working from home than average workers, especially in poorer countries. For example, while the authors' surveys suggest that one-third of American workers were working from home in 2021, the nationally representative American Community Survey finds that only 17.9 percent of work was at home during that year (US Census Bureau 2022). Second, at this moment labor seems scarce and employers are particularly prone to produce perquisites, including the option to work from home. Working from home makes a particularly tempting temporary perquisite because it can be readily removed as the labor market cools down and the preferences of employers become more important.

Third, basic economics implies that the owners of commercial property will cut rents rather than letting office space remain vacant for long periods of time. This equilibrium response should attract scruffier firms to formerly expensive real estate in cities like New York and San Francisco. Fourth, many of the advantages from collocated work remain, including the ability to share common infrastructure, the ability of employers to reduce distractions, and the ability to connect face-to-face. Many of these advantages are likely to become more, not less, important over time. Fifth, there appear to be dynamic losses from working from home that seem likely to become more apparent to both workers and firms over time, and this will bring people back to the office.

I am not confident about my more minimalist stance on working from home. There remains tremendous uncertainty, both about the course of technology and about the path of future pandemics. I am, however, quite confident that face-to-face contact is tremendously powerful and that it will play a central role in more of human productivity for the foreseeable future.

MEASURING WORKING FROM HOME IN THE UNITED STATES AND ACROSS THE WORLD Broad international evidence on the prevalence of, demand for, and expectations about the future of work from home is a fantastic contribution of this paper. Instead of extrapolating from the US experience, we can actually see how at least some part of the population is working from home everywhere. Moreover, the results seem broadly sensible and many of the findings seem in line with the predictions of a simple price-theoretic model of working from home.

Yet the most obvious limitation of this work, and indeed any internet-based survey, is the representativeness of the sample. As connecting in cyberspace is almost the defining feature of working from home today, we would be surprised if a survey delivered in cyberspace doesn’t overstate the amount of working from home. Even Google mobility data, which seem far more likely to be representative in the developing world than the authors’ surveys, surely suffer from some bias because of its dependence on internet-linked devices.

Following decades of (occasionally erroneous) practice, I am going to treat the surveys produced by the US Census Bureau and Bureau of Labor Statistics as ground truth. These public agencies have worked to generate representative samples for many decades. Their surveys are administered by phone and in face-to-face contact, and so they seem far less likely to miss the computer illiterate.[1] Unfortunately, these data sources do not give us anything like the regular updates on working from home provided by the authors' work or by the Google mobility data.

The Census Bureau began providing monthly data on working from home in May 2020, when supplemental questions on telework were added, and in particular respondents were asked "at any time in the LAST 4 WEEKS, did you telework or work at home for pay BECAUSE OF THE CORONAVIRUS PANDEMIC?"[2] This question became progressively less useful over time. In May 2020, it seems safe to assume that pretty much everyone who was working from home accepted that this was "because of the pandemic." Two years later, most of those who were working from home may have thought that the pandemic had little to do with the matter and that convenience or productivity caused them to work from home.

The Current Population Survey (CPS) reports that in May 2020, 35.4 percent of employed individuals were working from home "because of the pandemic." The educational skew in telework was enormous: whereas only 13.3 percent of employed Americans with a high school degree or less worked remotely, 59.6 percent of those with a college degree or more worked remotely.[3] According to this data source, the share working remotely (because of COVID-19) dropped to 21.2 percent of the employed population by October 2020, 11.6 percent by October 2021, and 5.6 percent by September 2022. Unfortunately, the only "fact" documented by this sharp downward trend is that people were no longer connecting working from home with the pandemic.

These monthly reports are the only time series made available by the government, but there are two other, presumably representative, samples available for the year 2021. Most importantly, the Census Bureau collected its standard American Community Survey (ACS), which attempts to provide representative data both for the United States as a whole and for larger geographic areas of the country (US Census Bureau 2022). The most relevant question is "How did this person usually get to work LAST WEEK?"

1. The Current Population Survey did stop face-to-face interviews during the pandemic, which may have altered the representativeness of the sample despite the best efforts of the Census Bureau (Ward and Edwards 2021).

2. US Bureau of Labor Statistics, "Measuring the Effects of the Coronavirus (COVID-19) Pandemic Using the Current Population Survey," https://www.bls.gov/covid19/measuring-the-effects-of-the-coronavirus-covid-19-pandemic-using-the-current-population-survey.htm.

3. US Bureau of Labor Statistics, "Effects of the Coronavirus COVID-19 Pandemic," table 1, https://www.bls.gov/cps/effects-of-the-coronavirus-covid-19-pandemic.htm.

Working from home is one of the options, which provides a measure of the share of the population who work from home more often than they go out to work. This survey question is, unfortunately, not well designed to measure the number of people who work from home one or two days per week.

According to this measure, 27.6 million Americans, or 17.9 percent of the employed workforce, typically worked from home in 2021 (US Census Bureau 2022). The same survey reported that 9 million Americans, or 5.7 percent of the employed, worked from home in 2019, and 5.9 million, or 4.3 percent of the employed, worked from home in 2010.[4] The 50 percent growth in the number of people working from home between 2010 and 2019 supports the view that this phenomenon had been growing significantly, if slowly, even before the pandemic.

The third public product which purports to provide a representative picture is the Bureau of Labor Statistics' Business Response Survey (BRS). This data source represents a supposedly "nationally representative survey of U.S. private sector businesses," but (like the ACS), this source only provides annual data on telework (Dalton and Groen 2022). The survey is taken between July and the end of September, and consequently there are results only for 2020 and 2021 at the time of writing.

In 2020, businesses were asked only if they offered telework or increased telework during the pandemic. The survey reported, for example, that 52.3 percent of all private sector establishments did not offer telework, and that 54 percent of all workers labored in establishments that had increased their level of telework since the pandemic.[5] Unfortunately, these numbers tell us little about the actual prevalence of teleworking across workers. More helpfully, the 2021 survey asked what share of the establishment's workers were remote either some or all of the time. As the survey also asks for the total number of employees, these data could be used to estimate the share of the American labor force that was either fully or partially remote between July and September of 2021 (Dalton and Groen 2022).

This survey finds that 12.6 percent of workers were fully remote, and another 9.2 percent were partially remote during that period. In some industries, like professional and business services and information, remote work

4. US Census Bureau, "American Community Survey: B08301, Means of Transportation," https://data.census.gov/table?q=B08&d=ACS+1-Year+Estimates+Detailed+Tables&tid=ACSDT1Y2010.B08301.

5. US Bureau of Labor Statistics, "Business Response Survey: BRS Tables," https://www.bls.gov/brs/data/tables/.

was ubiquitous, with 46.3 percent and 68 percent of workers, respectively, who are fully or partially remote. In other industries, such as manufacturing and accommodation and food services, remote work was rare with only 12.2 percent and 1.8 percent of workers fully or partially remote.

How do these numbers compare with the US Survey of Working Arrangements and Attitudes (SWAA), which is the model for the Global Survey of Work Arrangements (G-SWA) used in this paper? The SWAA asks, "Currently (this week) what is your work status?" and "working from home" is one of the available answers to this question. This question seems closest in spirit to the ACS's question about "usually" getting to work last week.[6]

We would certainly expect the number of people answering yes to this question to be smaller than the share answering yes to CPS's question about teleworking at any time in the last four weeks. However, 61.5 percent of the respondents to the SWAA reported working from home in May 2020, as opposed to 35.4 percent in the May 2020 CPS. In July, the SWAA working from home share had declined modestly to 51 percent, but the CPS share dropped to 26.4 percent. It is possible that some share of the discrepancy between the two measures reflects the "because of the pandemic" clause in the CPS question even during these early days. The alternative interpretation is that these data significantly overstated the share of Americans working from home during these months.

The Google mobility data create a third possible measure of working from home that is available at daily frequencies and fine spatial resolution. These data measure the change in the number of devices visiting particular locations, such as workplaces, relative to a comparable day of the week before the pandemic. These data will be biased if a nonrandom sample of the population use such devices, or if the prevalence of devices in the population is changing over time, which is particularly plausible in the poorer parts of the world. The data will capture declines in workplace visits, both because of telework and because of reductions in total employment.

During the week of April 27, 2020 (which includes May 1), the average number of workplace visits had declined by 47 percent relative to before the pandemic. This decline is substantially higher than the 35.4 percent CPS figure but lower than the 61 percent reported in the SWAA. However, the CPS also reports the 19.2 percent of workers in May 2020 who had lost their

6. The SWAA survey data and questionnaires can be accessed at WFH Research, www.WFHresearch.com.

jobs because of the pandemic.[7] If that number is added to the 35.4 percent, then the match to Google Community Mobility looks quite good. If that number is added to the 61 percent that the SWAA reports as working from home, then the decline in workplace visits should presumably be closer to 80 percent than 45 percent.

By the first week of July 2020, the Google mobility data report a workplace decline of 37 percent. The CPS reports 12 percent of the population having lost their job because of the pandemic, which suggests a total decline in workplace visits because of joblessness and telework of 38 percent. The CPS and Google Community Mobility again seem quite compatible with one another. The SWAA report of 51 percent working remotely continues to differ dramatically from the other two sources.

What about the results from the ACS and BRS? Over the year 2021, the SWAA reports that an average of 33.6 percent of respondents worked from home and that 32.7 percent of respondents worked from home over the July–September period. By contrast, the ACS reports that 17.9 percent of respondents worked from home during 2021. If one-half of the partially remote employees in the BRS would report themselves as primarily remote, then the equivalent figure for that survey is 17.2 percent. While both of these figures are quite similar, they each come to 52.3 percent of the SWAA figures.

In this case, the Google mobility data, which suggest roughly 30 percent declines in workplace visits during 2021, are much closer to the SWAA. Some of the gap with the ACS and BRS can be explained by lower employment, but for 2021, we are left with the disturbing possibility that the official products could both be wrong and both be underestimating the level of working from home. One possibility is that a large number of the 82.1 percent of ACS respondents are telecommuting one or two days per week. While that seems compatible with the experience many of us have in our own offices, it seems more difficult to reconcile with the BRS, which reported that only 9.2 percent of employees were working from home part of the time. It is also fails to reconcile the SWAA and the ACS, since both surveys focus on the predominant work experience "last week."

An alternative possibility is that Google Community Mobility may be overestimating the decline in workplace visits, either because of its sample of users or because of its definition of workplaces. It is possible that the set

7. US Bureau of Labor Statistics, "Labor Force Statistics from the Current Population Survey," https://www.bls.gov/cps/effects-of-the-coronavirus-covid-19-pandemic.htm; table 3, https://www.bls.gov/cps/covid19/covid19-table3-2020-05.xlsx.

of people who allow Google to track their location are disproportionately technology-oriented and so the data set is also disproportionately capturing technology-savvy, younger workers. A second possibility is that workplaces disproportionately mean offices, rather than restaurants, retail shops, and other places of work that involve face-to-face contact. Restaurants and retail have their own place category, and these show little evidence of any permanent decline in visits.

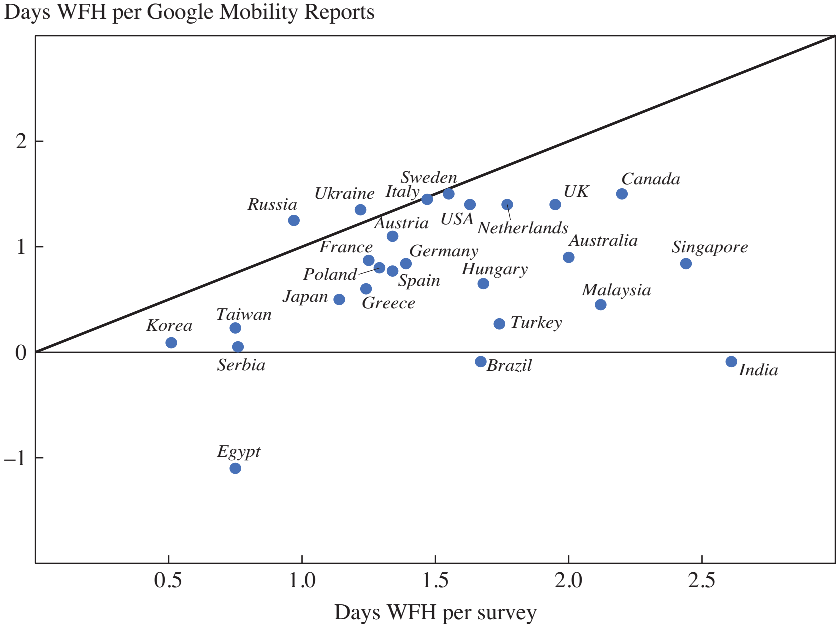
If the SWAA significantly overrepresents those who are comfortable with technology, then this may bias the estimated number of people working from home, but also other metrics, such as the estimated productivity impact of working from home. It seems quite possible that those people who like computers and are good at using the internet are more likely to answer internet-based surveys, work from home, and find working from home pleasant and productive. Consequently, it is difficult to know the representativeness of their results on the productivity benefits and desirability of working from home, although I have no doubt that millions were able to do their jobs well remotely and that almost all people like the option of working from home.

The problem seems potentially more severe in the G-SWA when the authors look at poorer countries. A simple way of examining the representativeness of their data is to compare the gender and education composition of their populations with the average education in the country, reported by Barro and Lee (2013), and the gender composition of the labor force. The share of men among their respondents ranges from 46 percent in Serbia to 53.6 percent in Russia, with Egypt as an extreme outlier with a male share of 76.2 percent (online appendix table A.3). Respondi is clearly aiming for gender balance. By contrast, the proportion of the labor force that is male, according to International Labor Organization data, ranges across these countries from 54.5 percent to 88.5 percent.[8] On average, the authors' sample is slightly more female than the countries as a whole, and excluding Egypt, there is little correlation between the gender balance of their samples and the gender balance of the labor force in the country as a whole.

An astounding 78 percent of the Indian respondents to the G-SWA have received graduate education; Barro and Lee (2013) report that in 2015 only 7.3 percent of Indians between age 25 and 64 have completed tertiary education. In Egypt, 86 percent of the surveyed population has tertiary or

8. World Bank, "Labor Force Participation Rate (% of Population)," https://genderdata.worldbank.org/indicators/sl-tlf-acti-zs/.

Figure 1. Working from Home according to Google's Community Mobility Reports



Sources: COVID-19 Community Mobility Reports, https://www.google.com/covid19/mobility/; and Bloom and others (2015).

graduate education in the survey; Barro and Lee (2013) report that 11 percent of Egyptians have completed tertiary schooling. The mismatch for education is less severe in the wealthy world.

The authors' quite reasonable procedure for dealing with this is to control for demographics and produce a country fixed effect, relative to the United States. They then add their country fixed effect to the US mean to produce their corrected measure. This procedure, for example, reduces India's reported current days working from home figure from 3.3 to 2.6 (online appendix table A.2 and figure 1). Ideally, this procedure tells us what India's number would look like if Indian education matched the US education levels reported in the sample, which includes 49 percent of individuals with tertiary or graduate education. Yet this thought experiment tells us little about what the actual Indian experience is likely to be going forward.

To compare these data with another source of information about working from home, I downloaded the Google mobility data for twenty-six countries (China is not covered) and calculated the average reduction in the number of workplace visits for the week of February 7, 2022.[9] To transform Google's estimated reduction in workplace visits into an estimated number of days working from home, I multiplied the Google change times –5, which would represent the number of average days working from home required to generate the reduction in workplace visits observed by Google.

The correlation between this measure and the authors' raw survey measure of working from home is shown in figure 1. The distance from the 45-degree line represents the discrepancy between their data and the Google mobility data.

In only two countries do the authors' data seem to underrepresent working from home relative to Google mobility: Russia and Ukraine. It seems possible that the war imminent at the time may explain the large reduction in workplace mobility in those countries.

In a number of countries, the authors' data match Google mobility quite well. Sweden and Italy, for example, show an almost perfect fit. The match in the United States is also quite good, which corresponds to my previous discussion of the congruence of the SWAA data and the Google mobility data. For most of the wealthier, Western countries, the results are quite similar.

Nonetheless, there are substantial discrepancies between the authors' data and much of the data outside the West. For example, in Brazil, Egypt, and India, Google workplace visits were actually higher in February 2022 than they were before the pandemic. This growth may reflect an increasing prevalence of cell phone ownership rather than the elimination of working from home; nonetheless, it does suggest that in these places working from home is an extremely elite phenomenon.

In most of the non-Western countries, including Australia, Korea, Japan, Malaysia, Singapore, and Turkey, the gap between the authors' data and the Google mobility data is significant. Most of these countries typically had fewer disruptions and fewer deaths from COVID-19. If we accept that the Google mobility data are more representative of the situation on the ground than the G-SWA, then this can be interpreted as providing support for the authors' core hypothesis: as these countries were shocked less by the disease, they remain trapped in the unfortunate equilibrium where people largely go to work. An alternative view is that Western countries were still working from home in February 2021 because of fear of the disease, which had largely disappeared from non-Western countries.

9. Google, "Covid-19 Community Mobility Reports," https://www.google.com/covid19/mobility/.

In table 1 in the paper, most of the basic facts on the gender gap, the role of children, and the complementarity between working from home and graduate education seem reasonable and seem likely to hold even for a broader sample. One way to interpret the negative effect of national income is that working from home is particularly appealing for people who are much richer than the society that they inhabit, partially because public services are so much worse than their private consumption levels. I don't know what to make of the impact of COVID-19 death rates, given that elite populations in poorer countries experienced COVID-19 in a very different way than the average resident of those countries. For example, Sheng and others (2022) report that "55% of Mumbai slums residents had antibodies to COVID-19, 3.2 times the seroprevalence in non-slum areas of the city according to a sero-survey done in July 2020" (abstract).

The G-SWA data are interesting and important, and no doubt show that many people have really liked working from home. Yet the selection of the samples bears closely on the question of whether the work-from-home revolution is likely to be permanent. If we think that the ACS and BSR figures of around 17.5 percent working from home in 2021 are more likely to be accurate than the SWAA figure of 33 percent working from home, then the empirical picture seems more ambiguous. For example, if that 17.5 percent were likely to move downward to 12.5 in a year or two, then the growth in working from home would seem far less like a permanent revolution than a continuation of the gradual increase in working from home that was already occurring prior to 2019.

MOVING TOWARD STEADY STATES IN COMMERCIAL REAL ESTATE AND LABOR MARKETS The authors write that "there are several reasons to think that WFH levels will ultimately settle at higher values than suggested by our survey data," including the "steady rise from January 2021 to June 2022 in the plans of American employers for WFH levels after the pandemic." In this section, I will argue that there are least two reasons why changes in the labor market and real estate market equilibria will push in the opposite direction. It is also worth noting that the SWAA-measured employers' post-pandemic plans for working from home have actually declined since June 2022. Moreover, we might wonder whether the employees who answer these questions actually know their employers' plans, especially since employers eager to retain their employees might be encouraging them to think that they will continue to have the option to work from home in perpetuity.

The authors estimate that "employees view the option to WFH two to three days per week as equal in value to 5 percent of earnings, on average," and some value the option for more than that amount. In many countries, employers have struggled to retain and attract workers since the start of the pandemic. In the United States, the pandemic in many ways seems to have been more of an adverse labor supply shock, sometimes called the "great resignation," than an adverse labor demand shock. Between October 2021 and October 2022, the US unemployment rate has been below 4.6 percent.[10] Basic economics suggests that firms will be more willing to offer perquisites when labor is difficult to retain and hire. As the US labor market reverts to more normal conditions, the labor market will slacken, and firms will presumably see less need to accommodate worker preferences for working from home.

Why would working from home be a particularly attractive tool for retaining and attracting labor during the current tight market? Increasing wage levels are hard to reverse. Bonuses are an alternative option, but even they create more of a precedent than simply continuing with a practice that was ubiquitous when the pandemic still raged. If a recession leads firms to have more bargaining leverage, then they will be able to change work-from-home conditions far more easily than they could change financial terms or conditions surrounding physical infrastructure. In many ways, working from home may be the easiest means of providing temporary benefits to workers during a tight labor market.

This argument means that current work-from-home levels could easily overstate, and perhaps significantly overstate, the longer-term level of working from home, but it does not suggest that working from home will disappear, even in a recession. In the longer term, workers will be richer and they will choose to take some of their earnings in the form of perquisites. One of those perquisites is likely to be working from home, which suggests that the pre-2019 trend away from the office will continue, even if there is an immediate decline in working from home in the aftermath of a recession.

The second equilibrium phenomenon related to working from home will occur in real estate markets. Over the course of the pandemic, Kastle has provided data on workplace occupancy across ten large metropolitan areas.[11] The data come from the use of security systems, which Kastle operates, and so change in occupancy reflects the change in the number of

10. US Bureau of Labor Statistics, "Databases, Tables and Calculators by Subject," https://data.bls.gov/timeseries/LNS14000000.

11. Kastle, "Kastle Back to Work Barometer," https://www.kastle.com/safety-wellness/getting-america-back-to-work/#workplace-barometer.

people swiping cards or fobs to enter large office buildings. Kastle reported an overall ten-city occupancy rate of 47.9 percent for October 19, 2022. The occupancy rate was higher in Houston (58.4 percent) and lower in San Francisco (41.2 percent). As the buildings that use a Kastle security system are unlikely to be representative, even in large downtown office markets, these data cannot inform us about the overall level of working from home. They do, however, imply that a lot of expensive commercial real estate is currently being underutilized relative to pre-pandemic norms. The standard logic of economics suggests that this should lead to a reduction in commercial rents, which should encourage occupancy by new tenants.

It is possible that some firms may reduce their usage of space without reducing their total consumption of commercial space even at existing prices. If a firm wants all of its employees in together on Tuesday, Wednesday, and Thursday, then it must continue to rent the same space even if everyone is working remotely on Monday and Friday. Yet it seems likely that many firms will try to reduce their physical footprint because of working from home. That reduction in demand strikes a relatively fixed supply of office space, and commercial rents should decline. The limited data that are available suggest that this is already starting to happen in some markets.

Firms that had been priced out of high-end office markets in 2019 and earlier may now think about moving into these markets. Lower commercial rents may encourage some entrepreneurship. If there is a substantial price effect, then any existing expectations about working from home will likely overstate the market-wide level of working from home. Almost assuredly, when individuals answer the G-SWA or SWAA surveys on plans for post-pandemic working from home, they are not thinking about how changes in commercial real estate costs may cause other firms to consider moving into downtown space.

The larger point of this section has been that there are good reasons to think that working from home may decline as the labor and office markets equilibrate. Firms will face less pressure to offer working from home as an option as workers become less scarce. Office rents will decline and induce more firms to opt to use those offices.

THE STATIC AND DYNAMIC COSTS OF WORKING FROM HOME In this penultimate section, I discuss the static and dynamic costs of working from home. I mention the static costs to suggest that it is not hard to figure out why many employers don't particularly like having a remote workforce, despite employees' preferences for at least having the option to go remote. I then discuss the dynamic costs of telecommuting that I suspect are less likely to be internalized, at least so far, by workers and firms.

Working in a common space has significant advantages historically: (1) workers can share fixed infrastructure, such as a textile loom; (2) managers can address worker incentives by monitoring behavior and limiting distractions; (3) workers and customers can meet in common spaces, such as a dining room; and (4) workers can collaborate in the short run and learn from one another in the long run. Of these four advantages, the first two are relatively untouched by advances in telecommuting. The last two advantages should be eroded by telecommuting technology, but it is unclear by how much.

Almost 18 million workers labor in the goods-producing sector of the US economy, and work from home seems likely to be limited in that sector because of the need to access factories, mines, and construction sites.[12] The BRS reported only 12.2 percent of manufacturing workers were doing at least some remote work in 2021, and the figure is lower for mining and construction (Dalton and Groen 2022). Similarly, working from home is ill-suited for wholesale and retail trade, leisure and hospitality, transportation and warehousing, and agriculture, although there will surely be back-office elements in these industries that can work from home. In the 2021 BRS, only wholesale trade had a significant amount of remote work (26.4 percent of jobs). Most of health care and social assistance also needs to be face-to-face, which is also corroborated by the BRS.

The key industries where working from home has been massive and which drive downtown office markets are financial services, professional and business services, and information, which collectively contained about one-fifth of America's labor force in 2021.[13] In these industries, there are no common infrastructure needs, and many elite workers are internally motivated and capable of using technology. It is less clear if secondary workers, such as lower-level administrators, in these industries will work hard when they are remote, but there is no question that many workers in these industries can do their jobs remotely. These three clusters, along with educational services, had levels of working from home that were substantially higher than the national average in the 2021 BRS. These are the industries in which the electronic innovations have had the largest impact.

Information technology has been available in these industries for decades, and yet these clusters are famous for their physical agglomerations. Information clustered in Silicon Valley; financial services clustered on Wall

12. FRED Economic Data, "All Employees, Goods-Producing," https://fred.stlouisfed.org/series/USGOOD.

13. US Bureau of Labor Statistics, "Employment Projections: Employment by Major Industry Sector," https://www.bls.gov/emp/tables/employment-by-major-industry-sector.htm.

Street (or midtown Manhattan). Despite the supposed death of distance, face-to-face contact remained a major part of life, whether for workers on trading floors or in the Googleplex.

One narrative for the surprising resilience of face-to-face contact in these industries is that technological change has done two things, which work in opposite directions. First and most obviously, information technology reduces the cost of long-range communication. Second and less obviously, the combination of technological change and globalization have significantly increased the returns to knowledge, information, and skill. If proximity enables the spread of knowledge, then the second force can outweigh the first, and that is one explanation for why high human capital cities have done well over the past forty years (Glaeser and others 2004).

According to this view, we should expect remote work to have risen significantly over the past two years, because innovation in long-distance connection has moved more quickly than any increase in the return to skill. Yet in the longer run, the old pattern may well reassert itself. In finance, much of the most important knowledge transfers occur at very high frequencies and so hybrid work (going in three days a week) is less plausible on a trading floor. In information services, the knowledge learned has a far lower frequency, which may well be compatible with working from home 40 percent of the time.

The dynamic benefits of face-to-face contact for knowledge creation are supported by the classic work-from-home paper by Bloom and others (2015) and by more recent work from Emanuel and Harrington (2021). Both papers find that the productivity of call center workers either rises or remains unchanged when those workers go remote. Both papers also find that the probability of being promoted drops by over 50 percent for the remote workers. These findings are compatible with the view that going remote shuts off part of the learning channel for both workers and their supervisors. Workers who disappear from the office completely will have little chance to learn from their colleagues or to shine in front of their supervisors. Workers who spend one day at home will still have plenty of chances to learn and to shine.

Other recent work supporting the learning-in-person channel comes from Morales-Arilla and Daboín (2021) and Yang and others (2022). Morales-Arilla and Daboín document the substantial and enduring decline in postings for jobs that could be done remotely during 2020 and 2021. This decline was not accompanied by a drop in employment. By contrast, both employment and job postings jumped back up in the summer of 2020 for jobs that had to be done in person. These findings are compatible with the view that companies did not want to onboard new workers who would be remote. This supports the hypothesis that working in person can be important for learning.

Yang and others (2022) examine the communications network within Microsoft after the firm went remote. They find that "firm-wide remote work caused the collaboration network of workers to become more static and siloed, with fewer bridges between disparate parts" and that "there was a decrease in synchronous communication and an increase in asynchronous communication." To the authors, these changes suggest that "these effects may make it harder for employees to acquire and share new information across the network" (abstract). Even if they are correct, however, these losses can probably be offset with workers coming back only 60 percent of the time.

I suspect that the dynamic losses from working from home will only appear over time, just like the losses that have already come from remote schooling. Many of the older workers will be fine with less learning, especially since they are the ones doing the teaching. The key question is whether employers will be willing to pay more to get the older workers to come to the office and enable the younger workers to learn from them. Classic human capital theory suggests that this will be the case if the young workers are learning firm-specific, not general, human capital. If the young workers are learning general human capital, then firms will only push the older workers to return if younger workers are willing to take a pay cut to have them around.

The social consequences of increased working from home. Increased working from home brings many benefits, especially for workers with small children. If the firm stays put, the primary impact of WFH will be to make longer commutes more tolerable, since the worker only needs to commute 80 percent of the time. The standard Alonso-Muth-Mills model of urban economics then predicts that successful metropolitan areas, like San Francisco, will get even larger and housing prices will drop more slowly with distance.

The new technologies will also make it easier for firms to relocate entirely even when there is no working from home. Moreover, the connection to downtowns will shrink further if WFH means that there are fewer providers of business services physically in those locations. This added mobility will make the fight to attract firms even more competitive and will punish cities that are not business friendly. Recent high-profile defections, such as the movement of Citadel from Chicago to Miami, suggest that the risks to older, colder cities are real, especially if crime rates begin to rise.

The cities of the developing world face many challenges, which may include high crime rates and contagious diseases, and always include terrible traffic congestion. Historically, these urban problems require the attention of urban elites who use their political clout to push for infrastructure, including aqueducts and sewers, that make cities healthier. The software engineers of Bengaluru are the best hope for an effective voting and lobbying bloc that can fight to improve that city's public services.

Yet when urban elites retreat, whether into suburbs or into their homes, they have less interest in fixing the city's larger problems. If the wealthy buy their own security teams, as they do in many Latin American cities, they have less interest in fighting for better policing for all. If WFH means that traffic becomes less of a problem for well-educated urban elites, then those elites have less interest in improving the roads of India's cities or in imposing congestion pricing. A reasonable guess is that technologies that enable rich urbanites in the developing world to rely less on common public services will only lead those services to become more problematic.

CONCLUSION This paper has significantly added to our stock of knowledge about working from home across the world. Even if the results in the poor world are highly nonrepresentative, they still suggest that WFH will remain the norm for a select group of privileged knowledge workers. In the wealthy world, Google mobility data largely confirm the authors' view that working from home is persisting. Even the minimalist view of working from home, articulated in this comment, accepts that millions of workers will labor at home a couple of days per week.

For most workers, the ability to work from home is an advantage, and I see few costs for the firms or their workers in the one day at home per week model. Yet that switch may have larger social costs which are not addressed by either this paper or my comment. Will working from home, or a related decline in business travel, significantly harm poorer workers who had provided services for downtown offices? Will working from home lead to even more of a disconnect between elite knowledge workers and the less fortunate, less educated workers who work in retail trade, leisure, and hospitality? The welfare consequences of working from home remain an important topic for future research, but it will be easier to assess those consequences in later years when we have more data.

REFERENCES FOR THE GLAESER COMMENT

Barro, Robert, and Jong-Wha Lee. 2013. "A New Data Set of Educational Attainment in the World, 1950–2010." *Journal of Development Economics* 104:184–98.

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying. 2015. "Does Working from Home Work? Evidence from a Chinese Experiment." *Quarterly Journal of Economics* 130, no. 1: 165–218.

Dalton, Michael, and Jeffrey A. Groen. 2022. "Telework during the COVID-19 Pandemic: Estimates Using the 2021 Business Response Survey." *Monthly Labor Review*, March. https://www.bls.gov/opub/mlr/2022/article/telework-during-the-covid-19-pandemic.htm.

Emanuel, Natalia, and Emma Harrington. 2021. "'Working' Remotely? Selection, Treatment, and the Market Provision of Remote Work." Working Paper. https://scholar.harvard.edu/eharrington/publications/working-remotely-selection-treatment-and-market-provision-remote-work.

Glaeser, Edward L., Albert Saiz, Gary Burtless, and William C. Strange. 2004. "The Rise of the Skilled City [with Comments]." *Brookings-Wharton Papers on Urban Affairs*, 47–105.

Morales-Arilla, José, and Carlos Daboín. 2021. "Remote Work Wanted? Evidence from Job Postings during COVID-19." Washington: Global Economy and Development, Brookings Institution. https://www.brookings.edu/wp-content/uploads/2021/07/Remote-work-wanted_July.pdf.

Sheng Jaymee, Anup Malani, Ashish Goel, and Purushotham Botla. 2022. "JUE Insights: Does Mobility Explain Why Slums Were Hit Harder by COVID-19 in Mumbai, India?" *Journal of Urban Economics* 127:103357.

US Census Bureau. 2022. "The Number of People Primarily Working from Home Tripled between 2019 and 2021." Press release, September 15. https://www.census.gov/newsroom/press-releases/2022/people-working-from-home.html#:~:text=SEPT.,by%20the%20U.S.%20Census%20Bureau.

Ward, Jason M., and Kathryn Anne Edwards. 2021. "CPS Nonresponse during the COVID-19 Pandemic: Explanations, Extent, and Effects." *Labour Economics* 72:102060.

Yang, Longqi, David Holtz, Sonia Jaffe, Siddharth Suri, Shilpi Sinha, Jeffrey Weston, and others. 2022. "The Effects of Remote Work on Collaboration among Information Workers." *Nature Human Behavior* 6:43–54.

GENERAL DISCUSSION John Abowd proposed that the authors subscribe to the American Association for Public Opinion Research Transparency Initiative so that the research community could benefit and better understand the team's survey methodologies.[1] Abowd argued that joining the initiative would set up meta-standards that guide key survey components such as sample recruitment, response rates, and sample comparability.

1. American Association for Public Opinion Research, "What Is the TI?," https://www-archive.aapor.org/Standards-Ethics/Transparency-Initiative/FAQs.aspx.

Robert Gordon seconded Katharine Abraham's discussion of how ambiguous it is to ask survey respondents how productive they were compared to their expectations. Gordon further argued that expectations can become even more ambiguous when confounded by the difficulty of knowing how long respondents work for. He cited a paper by Barrero, Bloom, and Davis, who found that roughly one-third of the time previously spent commuting was now spent at work.[2] This means that the total productivity measured in the survey may be a mix between actual higher productivity per hour and more hours of work in time that workers had previously spent commuting.

Gordon offered a more direct method of measuring productivity, which was to construct quarterly productivity data from the National Income and Product Accounts and the Bureau of Labor Statistics for particular industries. He reported that for service industries where people primarily work from home, such as finance and information, productivity went up at an annual rate of 3 percent in 2020. In the five quarters of 2021 and 2022, productivity increased even faster at an annual rate of 4 percent. Contact services decreased in productivity at an annual rate of about 2 percent in the same time period.[3]

Caroline Hoxby suggested more straightforward survey questions instead of the current question that asked about the worker's productivity relative to their pre-pandemic expectations. One question would ask for the number of hours spent working, including commuting; a separate question would ask about the productivity per hour prior to the pandemic and after.

Jonathan Wright echoed a similar sentiment when he asked the panel if there were studies that showed how work from home and productivity varies in jobs where output can be directly measured at high frequency, such as how many calls a worker takes in a call center, compared to jobs where it is harder to say what the day-to-day output has been.

Steven Davis pushed back on these comments by clarifying that the reason to include the question about productivity relative to expectations was to get at the particular mechanism of learning and revising priors that leads to re-optimizing work plans. He maintained that identifying that mechanism is difficult to do in other ways. With such a strong relationship

2. Jose Maria Barrero, Nicholas Bloom, and Steven J. Davis, "Why Working from Home Will Stick," working paper 28731 (Cambridge, Mass.: National Bureau of Economic Research, 2021), https://www.nber.org/system/files/working_papers/w28731/w28731.pdf.

3. Robert J. Gordon and Hassan Sayed, "A New Interpretation of Productivity Growth Dynamics in the Pre-pandemic and Pandemic Era U.S. Economy, 1950–2022," working paper 30267 (Cambridge, Mass.: National Bureau of Economic Research, 2022).

between the worker's assessment of their productivity surprise and the employer's plan for what the worker will do, Davis asserted that there is, in fact, a lot of information in those productivity surprises.

Davis acknowledged Abraham's discussion about primacy bias in the survey by noting that if it were present in the relationship between productivity surprises and planned work-from-home (WFH) days, it would attenuate the relationship, given the current response ordering for those questions. Davis claimed that there is mild evidence of primacy bias in their survey responses, but that the authors take the point and are moving to more use of randomized response options in future survey waves. To Abraham's point about survey responses being potentially biased because they reflect socially desirable outcomes, Davis said he was less worried and would leave it to the audience to judge whether their survey instrument tilts the responses one way or another.

Elaine Buckberg added to the social desirability issue by noting that responses might also vary across the business cycle. Responses during the current tight labor market with ample jobs may reflect this, and workers may become more willing to come to work in person or make location adjustments once the labor market softens. Buckberg also referred to a joint study by the Manufacturing Institute and Deloitte to highlight the point that the desire for flexibility is not just concentrated among white-collar workers but also among those who work hands-on in manufacturing jobs.[4]

Justin Wolfers emphasized that remote working has also allowed for more inclusivity. He reflected on past conferences of *Brookings Papers for Economic Activity* and the National Bureau of Economic Research being invitation-only and in-person events, while the transition to online screening has made them more accessible to others. Wolfers wondered if the same is also true among workplaces, with the expected primary beneficiaries being parents. He then questioned how confident we can be of current macroeconomic indicators with the immense shift to remote work.

Frederic Mishkin addressed the concern that working from home will decrease collaboration and innovation by referencing how workers in academia have been able to balance their remote work and flexibility with collaboration. Mishkin argued that one has less reason to be concerned if firms can learn how to accommodate individual schedules and coordinate particular on-site and off-site days. Hoxby also challenged the notion that firms need five days a week in the office in order to do spontaneous

4. Deloitte Insights, *Creating Pathways for Tomorrow's Workforce Today* (London: Deloitte Development LLC, 2021), https://www.themanufacturinginstitute.org/wp-content/uploads/2021/05/DI_ER-I-Beyond-reskilling-in-manufacturing-1.pdf.

collaboration. She argued that it is actually important to have a day or two away from other workers in order to finish projects and to spend the remaining days collaborating.

John Haltiwanger commented on how the spatial structure of the economy within a city might change when it comes to applications for new businesses. While applications for new businesses surged dramatically overall in 2020 and remained high throughout 2021 and 2022, the growth rate of new businesses during the pandemic was relatively low in areas such as Manhattan, relative to the surrounding counties in the New York metropolitan area.[5]

Gordon predicted that the cons of working from home would show up in the long run on downtown commercial real estate. He believed that as leases eventually come up for renewal, firms will decide to use less space, causing a collapse of commercial office construction and leading to a devastating effect on surrounding service businesses. Stijn Van Nieuwerburgh corroborated Gordon and stated that the number of newly signed leases for offices in some markets has fallen from 250 million per year to less than 100 million per year. Van Nieuwerburgh thought this impact would occur gradually: among all in-force leases as of the end of December 2019, only 38 percent came up for renewal in 2020 and 2021 combined, meaning there are still many firms that have yet to make decisions of whether to clear their office spaces.[6] Van Nieuwerburgh believed that the decline in property tax revenues from offices could potentially lead to an underfunding of mass transit and other public amenities. He then pondered how local decision makers can balance the tension between the local negative externalities created by remote work and the overall boost to productivity.

Hoxby pointed out that residential real estate and gentrification might also be affected if the extra day or two working remotely makes the home property further away from the city seem more appealing. Furthermore, stores in downtown areas might be adversely affected because, aside from more online shopping, individuals may now shop closer to home rather than coordinate it with going into the office.

Jason Furman was perplexed that his personal conversations with managers and business executives revealed completely opposite, negative opinions of working from home from what the authors presented. In trying

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

6. Arpit Gupta, Vrinda Mittal, and Stijn Van Nieuwerburgh, "Work from Home and the Office Real Estate Apocalypse," working paper, Social Science Research Network, November 26, 2022, figures 5 and 4, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4124698; data are from Compstak.

to understand the disconnect, he imagined that these managers might not see the same productivity gains or that they do not see the same willingness to take compensating differential pay cuts. Davis acknowledged that there is heterogeneity in how business people think about remote work, but he referenced evidence based on data from Lightcast (formerly Burning Glass) that show a sharp upward trend in job postings that allow for remote work since the summer of 2020.[7] Moreover, Davis said that SWAA data for the United States show that the initial employer resistance to working from home has gradually eroded and employer plans for work from home levels post-pandemic have drifted up since late 2020.

Betsey Stevenson remarked that while she was skeptical of the actual magnitudes in the willingness to pay to work from home estimate, she thought the paper did a better job at capturing differences between groups. She noted that understanding differentials is just as important because it could show how workers sort across communities and jobs and how that might have an impact on the gender wage gap. She appreciated the finding that people with the biggest productivity surprises are the most likely to keep working from home and claimed that it is evidence that businesses do experience learning shocks and correct their priors. To Gordon's point on directly measuring productivity, Stevenson added that another useful exercise is to compare the authors' productivity estimates to those by Fernald and Li, who examine the impact of COVID-19 on productivity and potential output.[8]

Gerald Cohen raised two questions. First, he asked whether the domestic outsourcing of workers—for instance, people living in Boise, Idaho, but working in San Francisco, California—would facilitate a trend to more international outsourcing, such as hiring workers who live in Bengaluru, India, but work for a San Francisco company. Second, Cohen inquired whether statistical agencies were collecting this information.

Davis concluded by encouraging researchers to access their data at the WFH Research website.

7. See Stephen Hansen, Peter John Lambert, Nick Bloom, Steven J. Davis, Raffaella Sadun, and Bledi Taska, "Remote Work across Jobs, Cities, and Countries" [slides], https://fbe.unimelb.edu.au/__data/assets/pdf_file/0020/4182320/remote_work_presentation_2.pdf; and Ethan Oldham and others, *Talent Playbook* (Boston: Lightcast, 2022), https://www.datocms-assets.com/62658/1663086344-lightcast-talent-playbook.pdf.

8. John Fernald and Huiyu Li, "The Impact of COVID on Productivity and Potential Output," in *Economic Policy Symposium Proceedings: Reassessing Constraints on the Economy and Policy* (Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City, 2022).

MAURICE OBSTFELD
University of California, Berkeley

HAONAN ZHOU
Princeton University

The Global Dollar Cycle

ABSTRACT The US dollar's nominal effective exchange rate closely tracks global financial conditions, which themselves show a cyclical pattern. Over that cycle, world asset prices, leverage, and capital flows move in concert with global growth, especially influencing the fortunes of emerging markets and developing economies (EMDEs). This paper documents that dollar appreciation shocks predict economic downturns in EMDEs and highlights policies countries could implement to dampen the effects of dollar fluctuations. Dollar appreciation shocks themselves are highly correlated not just with tighter US monetary policies but also with measures of US domestic and international dollar funding stress that themselves reflect global investors' risk appetite. After the initial market panic and upward dollar spike at the start of the COVID-19 pandemic, the dollar fell as global financial conditions eased; but the higher inflation that followed has induced central banks everywhere to tighten monetary policies more recently. The dollar has strengthened considerably since mid-2021 and a contractionary phase of the global financial cycle is now underway. Owing to increases in public- and business-sector debts during the pandemic, a strong dollar, higher interest rates, and slower economic growth will be challenging for EMDEs.

Since the late 1970s, cycles of US dollar appreciation have been accompanied by slower global economic growth, with the negative correlation most pronounced for emerging markets and developing economies (EMDEs). This time is no different. It may be surprising that this correlation

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 an interest in this paper.

Brookings Papers on Economic Activity, Fall 2022: 361–427 © 2023 The Brookings Institution.

has not weakened over the decades in light of the secularly declining economic weight of the United States on the production side of the world economy and the rising weight of the EMDEs. In 1992, the United States accounted for 19.6 percent of world GDP measured at purchasing power parity, versus a 42.3 percent share for EMDEs; by 2021 the US share had shrunk to 15.7 percent, whereas EMDEs had reached a 57.9 percent share of world output.[1] Nonetheless, fluctuations in the US dollar continue to play a key role worldwide and an especially powerful role in the fortunes of the less advanced economies. A fundamental reason is the explosive growth of global financial markets since the early 1990s and the dominant position of the US currency in those markets.

In this paper, we document the channels of the dollar's impact on EMDEs, building on recent research that seeks to trace and understand the international propagation of financial shocks. We emphasize how newer models of international finance have grown from earlier approaches in the face of the occasionally turbulent evolution of world capital markets. We also explore empirically the implications of those models for the US dollar's exchange rate. The paper is in four sections.

Section I makes three main points. First, in the fifty years since the emergence of the floating exchange rate system, the volume of international financial transactions has exploded compared with directly trade-related transactions. That expansion has brought a global financial cycle in world asset prices, leverage, and financial capital flows to the fore as a correlate of synchronized growth movements across countries. Second, as global financial markets have expanded in importance and scope, open-economy macro models have evolved to feature a more-detailed focus on financial markets along with the roles of risk aversion, market frictions, and investor sentiment. These models have yielded important insights on the international transmission of government policies and the factors behind exchange rate volatility. Third, even a half century after the advent of floating, the US dollar remains the world's dominant currency for asset markets as well as trade, making the nominal dollar exchange rate a reliably powerful concomitant of the global financial cycle. We document the dollar's strong negative correlation with key global real and financial variables, as well as its particular

1. IMF, "World Economic Outlook Database," for April 2022, https://www.imf.org/en/Publications/WEO/weo-database/2022/April, accessed August 15, 2022. The changes differ in magnitude but go in the same direction when market exchange rates are used to compare GDP shares. Using that metric, the US share drops from 25.7 percent to 23.8 percent between 1992 and 2021 while the EMDE share rises from 16.5 percent to 41.7 percent.

importance for emerging economies, and list features of EMDEs that help to explain this correlation.[2]

In section II, we illustrate the pervasive influence of dollar shocks on EMDEs by tracking their dynamic relation to a range of quantity, price, and financial variables. We argue that with appropriate econometric controls, the dollar's weighted nominal exchange rate against other advanced economies can be viewed as an external predictor of macro developments in EMDEs. Using a panel local projections (LP) framework applied to a set of twenty-six EMDEs over 1999–2019, we document that dollar appreciation shocks predict declines in output, consumption, investment, and government spending. Accompanying these developments are a decline in the traded-goods sector, a depreciation of the local currency against the dollar, a fall in the terms of trade (that is, a rise in the price of imports relative to exports), a decline in domestic credit, losses in equity markets, and a widening of the sovereign borrowing spread for foreign currency loans. These adverse correlates of dollar appreciation shocks are more pronounced for countries that peg their exchange rates, that have not adopted inflation-targeting monetary frameworks, and that have high levels of external liabilities denominated in US dollars. One policy inference consistent with these findings is that more-flexible exchange rate regimes do not shut out the global financial cycle, but they are indeed helpful in buffering external financial shocks and can do so most effectively when supported by relatively high inflation credibility at the central bank and relatively low external dollarization.

To understand better the US dollar's powerful influence over EMDEs' macroeconomic and financial conditions, we next seek to identify factors that drive the shock variable in our local projections, the dollar's exchange rate against other advanced economies. Section III reports the results of that investigation over the 1999–2021 sample period. US monetary policy (proxied by the change in short-term US Treasury rates) is an influential correlate of dollar movements; so are long-term Treasury rates, which have played an especially important role during the Federal Reserve's large-scale asset purchases of the zero lower bound period, but not just then.

Recent literature on exchange rate determination, surveyed below, has also found an important role for investors' perceptions of the safety and

2. We follow the literature in our focus on the nominal dollar exchange rate because it is that variable that adjusts in the short run to financial shocks. The real exchange rate is more relevant for resource allocation, but in environments with moderate inflation, changes in real and nominal rates are highly correlated.

liquidity of US Treasury assets, proxied by deviations from covered interest arbitrage in government bond markets. This factor creates a potent interaction between the global financial cycle and the dollar, because in "risk-off" episodes where global risk appetite declines, investors' flight to safe assets simultaneously raises the foreign currency price of dollars and constrains the lending of financial intermediaries. Like other recent authors, we find a prominent role for the relative US Treasury "convenience yield" in section III, and we make a case that this attribute of Treasury obligations depends in large part on the perceived safety and liquidity conferred by their dollar denomination. A direct indicator of low investor risk appetite, the excess bond premium (EBP) proposed by Gilchrist and Zakrajšek (2012), turns out to be the most reliably influential correlate of dollar movements in our estimates. An examination of the EBP's influence on EMDEs in the LP framework of section II implies that dollar movements driven primarily by changes in the EBP predict especially large and persistent negative effects.

Our concluding section IV places the current troubled global economic landscape in the context of the global dollar cycle. High inflation driven in part by a sharp recovery from the COVID-19 recession sparked a monetary tightening cycle across major central banks. In response, the world economy moved from an expansionary phase of the global dollar cycle following the initial COVID-19 shock in the first half of 2020 to a contractionary phase now. The Federal Reserve has been among the most aggressive (if not early) tighteners, and the dollar has appreciated sharply since mid-2021. Determined disinflation by the Federal Reserve and continued dollar appreciation could lead to more intense debt troubles for a range of EMDEs. Indeed, danger signals are flashing. On the other hand, if the Federal Reserve fails to get a handle on US inflation, that would be disruptive in the longer term. Among the consequences, the dollar's status as the premier global currency could come under threat, reinforcing other disintegrative trends and risks.

I. The Dollar and the World Economy: Evolving Linkages and Models

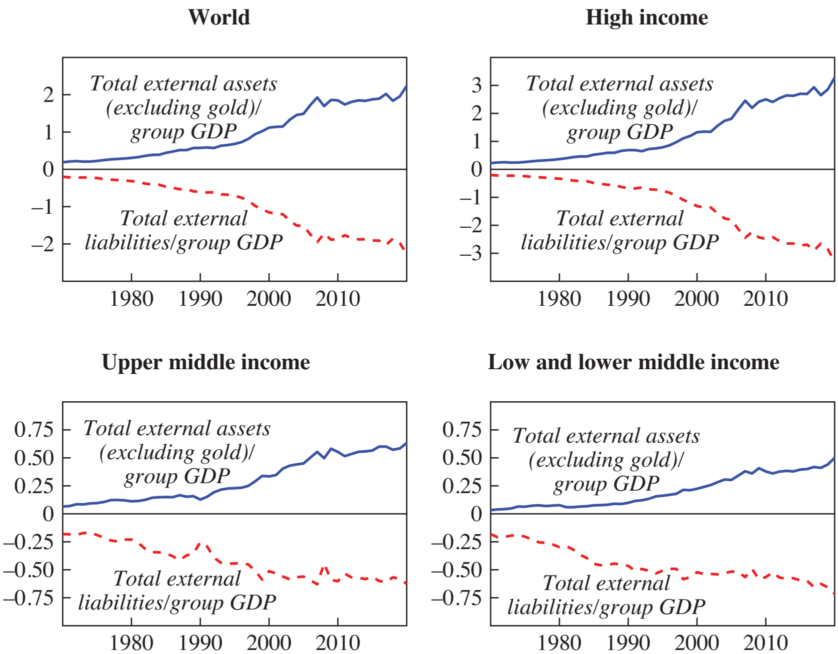
The modern system of floating exchange rates was born in March 1973, just short of fifty years ago. Having faced a long period of intense speculative pressure in foreign exchange markets, Japan and a large group of European countries suspended nearly three decades of postwar practice in that month and announced they would no longer peg their currencies to the US dollar. In the subsequent half century, what initially looked like a temporary retreat from the dollar-centric Bretton Woods system became permanent, and by the turn of the millennium, many EMDEs had embraced considerable exchange rate flexibility as well. These developments took place in a global environment of supply shocks and high inflation and were in part motivated by countries' desire to sever links with the dollar that made it hard to manage domestic macroeconomic policy independently. Yet, despite that intention, the dollar has remained central to the functioning of the international monetary and financial system, as has US monetary policy. The system has evolved considerably, however, and with it, the ways in which US policies and the dollar have an impact on the rest of the world.

The most notable change has been a spectacular growth in international financial positions and flows, facilitated by the rapid deepening of national financial markets and their cross-border linkages. Due to this growth, the way economic shocks are propagated through the world economy has changed. One important change following the initial years of floating is that US macroeconomic policies have increasingly come to affect other countries through financial channels, even countries with exchange rates that are flexible against the dollar. Another change is the greater scope for global financial market shocks to buffet the dollar, with spillback effects outside the United States, particularly in EMDEs. In this section we survey key indicators of the changes in global capital markets, important co-movements between global macro-variables and the dollar, and ways in which open-economy theories have progressed to address these facts.

I.A. Trends in Global Capital Markets

The end of the industrial countries' fixed exchange rates in the early 1970s set off a process of wide-ranging financial account liberalization. Without some degree of restriction on cross-border financial flows, the Bretton Woods system would likely have fallen victim to speculation even before the early 1970s. The adoption of floating, however, eased balance of payments constraints and allowed countries to direct monetary policy toward domestic rather than external goals, while simultaneously freeing up cross-border payments. That countries suddenly had the option to liberalize international financial flows does not fully explain why they chose that path. The political and economic factors pushing in that direction were sufficiently powerful and widespread, however, that by the mid-1990s the richer economies were approaching an unprecedented degree of financial integration while many

Figure 1. External Asset and Liability Ratios to GDP across Country Groups, 1970–2020



Source: Data from Lane and Milesi-Ferretti (2018), updated through 2020, https://www.brookings.edu/research/the-external-wealth-of-nations-database/.

Note: Income groupings for this figure are based on the 2019 World Bank classification. These data exclude small offshore financial centers in the Caribbean and Channel Islands.

emerging markets embarked on more limited, but still substantial, liberalization programs.[3]

One indicator of a country's global financial integration is the level of external assets and liabilities that it holds, measured as a ratio to GDP. Figure 1 plots these data for the world economy as a whole, as well as for three groups of countries: high-income, upper-middle income, and lower-middle plus low-income economies. These ratios increased markedly after the early 1970s, accelerating upward around the mid-1990s before continuing their advance at a slower rate after the global financial crisis of 2007–2009.

3. For historical perspectives on the evolution of the global capital market emphasizing economic and political drivers, see Obstfeld and Taylor (2017) and Obstfeld (2021).

Several facts stand out. For the advanced, high-income economies, external positions now exceed three times GDP on a weighted-average basis. In some cases, such as that of the United States, external positions are levered and subject to substantial currency mismatch, meaning that movements in equity prices, bond prices, and exchange rates—sometimes driven by waves in global investor sentiment—can effect sizable transfers of wealth from or to foreigners.

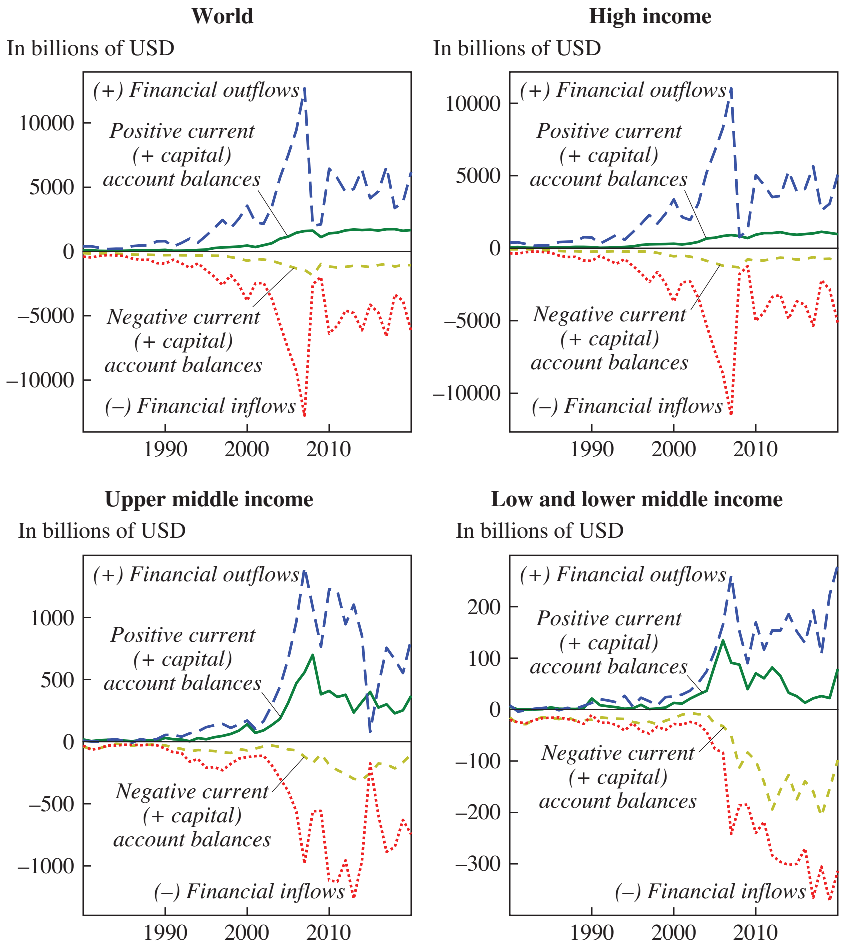
The two EMDE income groups hold broadly similar levels of external assets and liabilities, but lower-income countries hold fewer external assets and more liabilities, making many of them substantial net foreign debtors. If we measure average financial integration by external asset ratios, EMDEs are now where the high-income countries were around the late 1980s. Given more market and institutional fragility in many of these countries, however, increasing financial openness has brought greater vulnerability to capital market disturbances—as Calvo, Leiderman, and Reinhart (1996) highlighted and as we discuss further below. Much debt of low-income countries is owed to official creditors, of which China is now the biggest, and some official debts carry concessional terms.[4] But lower-income "frontier markets" are quite exposed to global financial shifts.

Short-term movements in exchange rates are driven by asset demand and supply changes that are reflected in financial account balance of payments flows. The greater importance of the financial account for exchange rate determination today owes to the huge volume of two-way traffic through foreign exchange markets to finance asset transactions, compared with the much more modest flows that would be the minimum necessary to finance current account imbalances alone.

Figure 2 offers one way to visualize the evolution in the external financing landscape. For the same groupings as in figure 1, figure 2 shows separately the sum of the included countries' current plus capital account surpluses and deficits—preponderantly balances of trade in goods, services, and investment income. The figure also shows separately global financial (often called capital) inflows, which are national residents' net incurrence of liabilities to foreign residents, and financial (or capital) outflows, which are national residents' net acquisition of claims on foreign residents. In principle, countries could finance their current account deficits with financial inflows just equal to those deficits (assuming no financial outflows) and dispose of their current

4. See Horn, Reinhart, and Trebesch (2019), whose estimates suggest that the size of China's official lending surpasses that of important multilateral institutions such as the World Bank and the IMF.

Figure 2. Global Current Account Imbalances and Financial Flows across Country Groups, 1980–2020



Source: International Monetary Fund, Balance of Payments Statistics.

Note: Income groupings for this figure are based on the 2019 World Bank classification. These data exclude small offshore financial centers in the Caribbean and Channel Islands.

account surpluses via financial outflows just equal to those surpluses.[5] As the figure shows, however, the volumes of two-way capital flows are much higher than that. Over the past decade, global capital inflows and outflows have been around $5 trillion annually, while global current account imbalances have been a small fraction of that. The same pattern holds even for the richer EMDEs.[6] Financial flows ballooned to extreme levels everywhere before the global financial crisis, receding sharply as the crisis unfolded.

These high volumes of financial flows provide a potent channel for external disturbances to have an impact on domestic asset markets as well as the real economy. A rise in world demand for a country's assets, for example, will result in financial inflows as well as currency appreciation and higher asset prices. These price changes will reduce the current account balance over time, but more quickly, they act to moderate the initial incipient financial inflow and induce a financial outflow owing to the lower expected return on domestic assets. In the process, those whose appetite for the target country's assets has risen end up holding more of them, while those domestic or foreign residents who part with those assets end up holding more foreign bonds, loans, and equities. Notwithstanding ex post financial account credits and debits that are largely offsetting, the process is far from neutral, as it has an impact on net exports, domestic aggregate demand, inflation, and financial conditions.

I.B. Global Cycles and the Dollar

Research following the global financial crisis has documented that the world economy is subject to synchronized cycles in asset prices, leverage, and capital flows. Financial cycles are driven in part by US developments, including Federal Reserve monetary policy, but also have an important global component that channels actions by major non-US central banks.

5. In principle, global current account surpluses should equal global deficits and global financial inflows should equal global outflows. Errors and omissions in balance of payments data, sometimes large, mean that these equalities do not hold exactly in practice. Financial flows to upper-middle-income countries were supported during the early 2010s by advanced economy central banks' large-scale asset purchases, but fell sharply in 2015–2016 in the face of turmoil in China's equity and currency markets.

6. In addition, while financial inflows and outflows as reported in balance of payments statistics are often referred to as gross capital flows (the net balance of financial outflows less inflows being the current account balance), they are net measures. Financial inflows are foreign residents' purchases less sales of domestic assets, while financial outflows are domestic residents' purchases less sales of foreign assets. The true gross transaction levels are big multiples even of the gross flows shown in figure 2. For example, see the discussion of the United States' international financial transactions in Obstfeld (2022).

The nominal exchange rate of the dollar is a prominent correlate of global financial conditions, with a stronger dollar implying increased financial stringency globally.[7] In EMDEs where there are significant private or public dollar liabilities, a stronger dollar tends to raise those liabilities' values, immediately impairing balance sheets and tightening financial and fiscal conditions. More than 80 percent of emerging markets' overall external debt liabilities are denominated in foreign currency, mostly US dollars (Financial Stability Board 2022), and in some countries, internal currency mismatch creates another potential fault line.[8] Not only does a stronger dollar itself lead to tighter financial conditions by weakening debtor balance sheets, heightened risk aversion in world markets tends to appreciate the dollar as investors everywhere seek safety, implying another channel of negative correlation between dollar strength and EMDE macroeconomic performance. Episodes of high global liquidity are associated with a weak dollar and lead to capital inflows and credit expansion in EMDEs, but a prior buildup of vulnerabilities can crystallize abruptly when the global financial cycle turns and the dollar strengthens.[9]

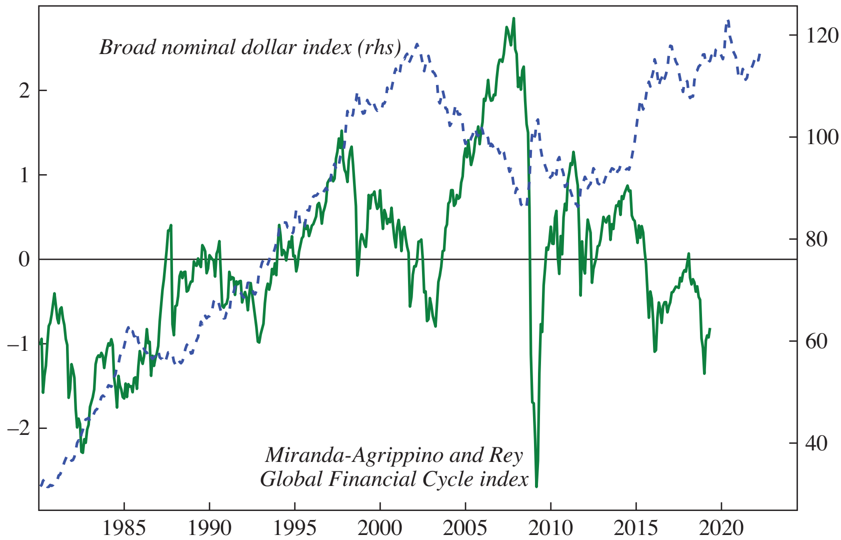
Figure 3 shows the relationship between monthly levels of the nominal effective US dollar exchange rate and the global financial cycle index constructed by Miranda-Agrippino and Rey (2020), as extended and updated by Miranda-Agrippino, Nenova and Rey (2020). Their index is defined as the common global factor from a dynamic factor model of equity, corporate bond, and commodity prices from markets in North America, Latin America,

7. On the global financial cycle, see Rey (2013) and the recent survey by Miranda-Agrippino and Rey (2022). Both the cycle and the dollar's central role were highlighted by Bruno and Shin (2015a, 2015b) and Shin (2020), and have been explored in subsequent work by these authors along with others. Important contributions by Reinhart and Reinhart (2009) and Forbes and Warnock (2012) documented the cyclical behavior of international capital flows, which is also evident in figure 2. Jordà and others (2019) offer evidence of a global financial cycle among seventeen advanced economies over the past century and a half. They document that its intensity has been historically high since around 1990.

8. The Financial Stability Board estimate of external foreign currency debt liabilities does not cover China. However, the net external US dollar debt exposures of China's banks and nonfinancial firms are large and growing, as the Committee on the Global Financial System (2020) and Kodres, Shen, and Duffie (2022) document.

9. The procyclicality of capital flows to EMDEs has risen in recent years as nonbank lenders, notably investment funds, have come to play a bigger role compared with banks (Financial Stability Board 2022). While more sovereign issuance in domestic currencies has mitigated the classic "original sin" fiscal vulnerability due to dollar issuance, it can promote capital flow volatility because advanced country investors in sovereign bonds are exposed to currency risk in addition to duration risk when advanced country interest rates rise and induce rises in EMDE rates. Carstens and Shin (2019) characterize this interplay as "original sin redux." EMDE corporates continue to borrow extensively in US dollars.

Figure 3. Broad Nominal Dollar Index and Miranda-Agrippino and Rey Global Financial Cycle Index



Sources: Miranda-Agrippino, Nenova, and Rey (2020); Federal Reserve H.10 release (FRED ticker DTWEXBGS).

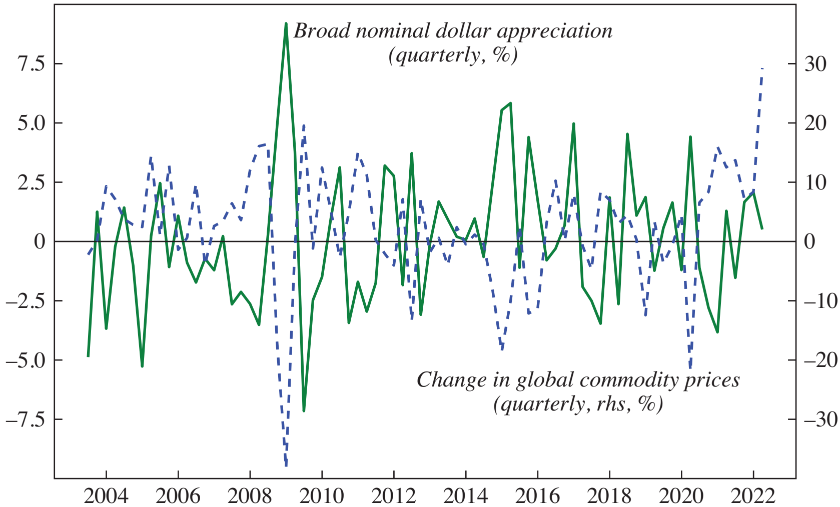
Note: The underlying currency weights are based on goods and services trade and are available at https://www.federalreserve.gov/releases/h10/weights/default.htm. The dollar index prior to 2006 is provided by von Beschwitz, Collins, and Datta (2019), the currency weights of which incorporate estimated services trade data.

Europe, and the Asia-Pacific region, including Australia. The correlation over the period since 2000 is quite negative, at –0.54. In the present millennium, tighter financial conditions have accompanied a stronger dollar.[10] Davis, Valente, and van Wincoop (2021) and Miranda-Agrippino and Rey (2022) show that common global factors in gross capital flows move closely with asset price factors.

Part of the mechanism underlying the negative correlation in figure 3 is a strong negative relationship between the dollar and global commodity prices, illustrated in figure 4. The correlation coefficient between the monthly

10. This levels relationship appears to be a medium-frequency one: the correlations between monthly changes are close to zero over the entire period in both the pre- and post-2000 subsamples. Over the entire sample period starting in 1980, the simple correlation coefficient between the levels of the two monthly series is positive at 0.47; and over the subperiod ending in 2000, it rises to a very high 0.79. These estimates could be misleading, however, because the coverage of the Miranda-Agrippino, Nenova, and Rey (2020) update in terms of both countries and assets is more limited before the late 1990s.

Figure 4. The Dollar and Dollar Commodity Prices



Sources: International Monetary Fund, Primary Commodity Prices; Federal Reserve H.10 release (FRED ticker DTWEXBGS).

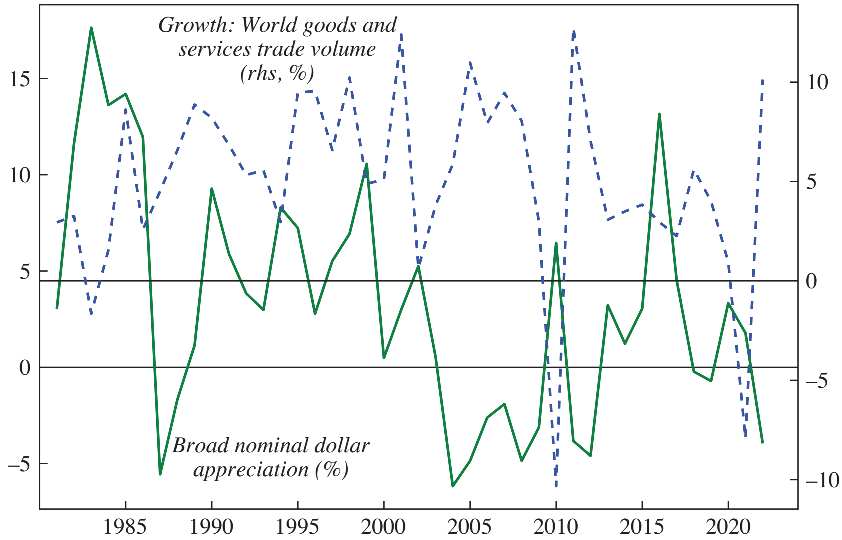
Note: The underlying currency weights are based on goods and services trade and are available at https://www.federalreserve.gov/releases/h10/weights/default.htm. The dollar index prior to 2006 is provided by von Beschwitz, Collins, and Datta (2019), the currency weights of which incorporate estimated services trade data.

changes is –0.57 over the period from February 2003 to April 2022. Observe the difference in scales between the left-hand vertical axis measuring dollar movements and the right-hand axis measuring commodity price movements. A 1 percent appreciation of the dollar is associated with a much larger percentage fall in average global commodity prices. Thus, dollar commodity prices fall in real terms when the dollar strengthens. In itself, this change generally hurts commodity exporters among the EMDEs while benefiting importers, but it is not the only implication for these countries of a stronger dollar.[11]

One implication, as figure 5 shows, is that the growth in world trade volume is strongly negatively correlated with changes in the dollar's strength.

11. Obstfeld (2022) discusses the dollar–commodity price link in more detail. See also Druck, Magud, and Mariscal (2018). The IMF index in figure 4 is an average over many commodities that can move idiosyncratically. For example, dollar appreciation in 2022 has been driven partly by high oil and agricultural prices that have pushed up inflation and elicited contractionary central bank responses. Yet, as expectations of a recession have risen, other commodity prices (such as industrial metal prices) have fallen.

Figure 5. Dollar Appreciation and Growth in World Trade in Goods and Services



Sources: International Monetary Fund; Federal Reserve H.10 release (FRED ticker DTWEXBGS).

Note: The underlying currency weights are based on goods and services trade and are available at https://www.federalreserve.gov/releases/h10/weights/default.htm. The dollar index prior to 2006 is provided by von Beschwitz, Collins, and Datta (2019), the currency weights of which incorporate estimated services trade data.

Partly this results simply from the importance of commodities in world trade—when their real prices fall, measured world trade volume contracts—but there are several other important channels at work, including financial channels. One is the key importance of trade in investment goods, with world investment being strongly negatively correlated with the dollar.[12] Table 1 documents the negative year-by-year correlations of the dollar with world trade and investment—and their increased absolute size—after the year 2000. Given these patterns in the data, it is not surprising that dollar strength is also negatively correlated with growth in advanced economies and in EMDEs, as table 1 and figure 6 show. EMDE economic fortunes are even more tightly linked to the dollar than are those of the advanced economies. Financial as well as trade channels are at work for both sets of

12. The International Monetary Fund (2016) documents the link between global trade volume and investment. For further discussion of dollar-trade causation channels, see Bruno, Kim, and Shin (2018), Bruno and Shin (2021), and Obstfeld (2022).

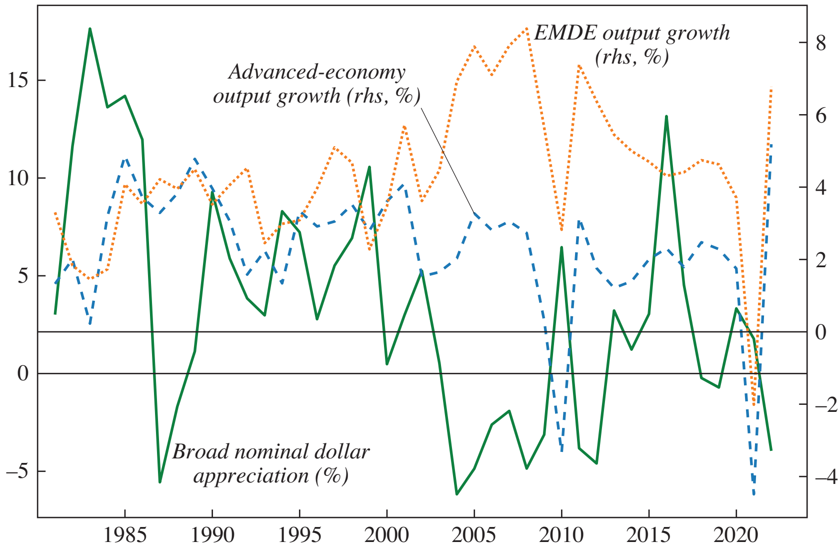
Table 1. Dollar Appreciation and Global Aggregates

Correlation with	*1980–2021*	*1980–2000*	*2001–2021*
World trade volume growth	–0.32	–0.39	–0.61
Growth in world investment/GDP share	–0.45	–0.32	–0.58
Advanced economy output growth	–0.05	–0.24	–0.36
EMDE output growth	–0.63	–0.56	–0.59

Sources: World Economic Outlook Database, April 2022; Federal Reserve H.10 release.

Note: Exchange rates are year averages of the broad dollar nominal exchange rate from the Federal Reserve H.10 release. The underlying currency weights are based on goods and services trade and are available at https://www.federalreserve.gov/releases/h10/weights/default.htm. Pre-2006 currency weights incorporate estimated services trade data; see von Beschwitz, Collins, and Datta (2019) for details. The data series for the change in world investment begins in 1981. The numbers reported are simple correlation coefficients of percentage changes in the exchange rate index and a global aggregate growth rate.

Figure 6. The Dollar and GDP Growth in Advanced Economies and EMDEs



Sources: International Monetary Fund; Federal Reserve H.10 release (FRED ticker DTWEXBGS).

Note: The underlying currency weights are based on goods and services trade and are available at https://www.federalreserve.gov/releases/h10/weights/default.htm. The dollar index prior to 2006 is provided by von Beschwitz, Collins, and Datta (2019), the currency weights of which incorporate estimated services trade data.

countries, and the relative importance of these channels has changed over time with the growth, scope, and reach of international financial markets.

I.C. Financial Market Experience and Exchange Rates

Early macroeconomic models of policy transmission under floating exchange rates focused on induced changes in the current account balance, which largely determined whether policies would be transmitted positively or negatively abroad. An expansionary monetary policy, for example, would raise output and therefore spending on imports, imparting a positive stimulus abroad, whereas the accompanying currency depreciation might shift domestic demand away from imports while raising exports, imparting a negative impulse. In these models, the net effect on foreign aggregate demand would be positive if the expanding country suffered a reduction in its current account balance, but negative if the current account balance improved. Capital flows played an entirely supporting role, passively financing any current account imbalance at a global interest rate equalized to the domestic rate (when reckoned in a common currency) through a risk-neutral uncovered interest rate parity (UIP) condition. To the extent that policies by the United States played any unique role, it was due to the country's size—its share of global GDP—which gave its policies the power to affect foreign rates of interest.

While the preceding channels have remained important, they offer an increasingly incomplete picture of either policy transmission or exchange-rate determination today. A half century after the move to floating, gross capital flows have expanded far beyond the needs of trade finance, and exchange rates must equilibrate these financial flows in the face of potentially large shifts in investor preferences and global asset supplies. Attention has therefore shifted to more-detailed accounts of the structure of international financial markets and the determinants of capital flows, along with the possibility that financial account drivers of exchange rates could appear dominant over short- and even medium-term horizons. The need to update exchange rate theories became more apparent after the global financial crisis. Since the crisis, frictions have become more salient in a range of financial markets, including international money markets, due to new financial regulations and changing business models.[13] The implications are especially

13. Early on, Dornbusch (1976) highlighted how exchange rates could react disproportionately to money supply shocks in models with sticky output prices, "overshooting" long-run positions even when investors have rational expectations and UIP holds. More recent models posit a role for possibly hard-to-observe financial market shocks, amplified by market frictions (Itskhoki and Mukhin 2021).

important for EMDEs, where the shocks to global financial markets collide with shallower and more brittle financial systems, institutions, and policy frameworks.

An important strand of theorizing from the 1970s and 1980s, recently revived, is the portfolio balance approach to capital flows and exchange rate determination. This approach views demands in international asset markets as reflecting optimizing choices by risk-averse investors, following the work of James Tobin.[14] UIP does not generally hold in these models, and uncovered interest arbitrage among currencies can offer positive or negative expected returns that depend on the covariance of returns with an appropriate stochastic discount factor (a risk premium). More recent models combine risk-averse investors with segmented financial markets where specialized traders operate. As in the main model of Gabaix and Maggiori (2015), departures from UIP can emerge even under risk neutrality if incentive constraints limit financial intermediaries' balance sheet sizes and thereby create limits to risk-neutral arbitrage. However, these models become even richer with risk-averse investors (Gabaix and Maggiori 2015; Itskhoki and Mukhin 2021). Another rationale for departures from UIP is based on the idea that bonds denominated in different currencies, and issued by different borrowers, may offer different degrees of liquidity. That additional "convenience yield" can compensate holders to some degree for a lower pecuniary return on the bond. Several studies have argued that US Treasury liabilities offer especially high convenience yields.[15]

A common theme in these models is that asset-demand functions are downward sloping: wealth owners will willingly absorb more of a particular bond onto their balance sheets only if its price falls, that is, if its expected yield rises. Downward sloping demand can be motivated by risk aversion, by the need for a bond's excess return to rise to compete for scarce balance sheet space, or by marginal convenience yields that diminish as the supply of a particular bond rises. Unlike in the UIP world, where bond demands are infinitely elastic, however, these models open the door to a

14. The approach was discussed in the pages of *Brookings Papers on Economic Activity* by Branson (1970), Kouri and Braga de Macedo (1978), and Dornbusch (1980), among others.

15. See, for example, Canzoneri and others (2008), Krishnamurthy and Vissing-Jorgensen (2012), Nagel (2016), and Del Negro and others (2017). Du and Schreger (2022) and Maggiori (2022) provide recent surveys of models with financial market imperfections. In these models, global risk-off episodes propagate through various channels, for example, increasing demand for asset safety and liquidity or, even in models where investors are risk neutral, constricting leverage due to tighter value-at-risk constraints (Adrian and Shin 2014). These different mechanisms may call for different policy responses to economic or financial shocks.

rich array of additional asset market shocks: to investors' risk aversion, to their appetite for safe assets or liquidity, to the stringency of financial constraints, to relative supplies of bonds in different currencies, or simply to non-optimizing behavior. Some of these shocks are driven by monetary policy, but they can arise independently of monetary policy or other central bank actions, and importantly, some appear to be major drivers of exchange rates.[16] A challenge for empirical work is to find measurable counterparts of these financial shocks.

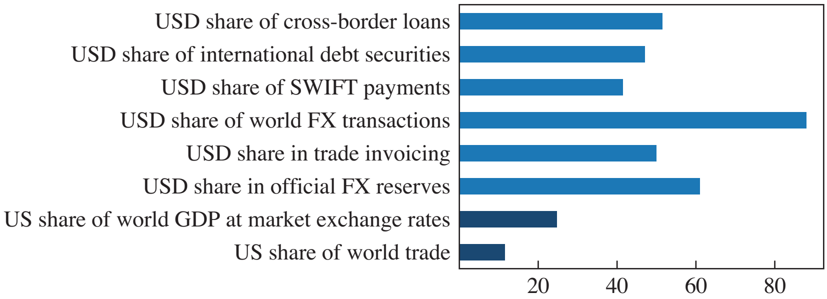
Although financial shocks need not be driven by monetary policy, monetary policy can affect financial conditions in ways that propagate internationally. Ammer and others (2016) find that US monetary policy tightening transmits abroad primarily through a financial channel—long-term US interest rates rise with direct spillover effects on foreign long-term rates. The resulting contractionary impact on foreign activity is the main net effect of US policy, as the impact on the US current account balance is minimal.[17] Monetary policies may also spill abroad by other effects on financial conditions, for example, through interrelated effects on investor expectations, balance sheet constraints, leverage, and risk aversion. US monetary policy is especially powerful in this regard, as documented by Miranda-Agrippino and Rey (2022), among others. Kalemli-Özcan (2019) argues that hikes in the federal funds rate lower the risk tolerance of global investors (the risk-taking channel of monetary policy), with particularly strong effects on capital flows, credit spreads, and sovereign borrowing premia in EMDEs.

The special importance for the world of US policies and financial conditions is hard to rationalize in traditional models, other than through the United States' global GDP weight, an attribute broadly shared by the euro area and China. However, the US footprint in financial markets is proportionally much larger than its GDP weight, and its financial markets are the deepest anywhere. As of 2021, for example, US equity markets accounted for over 40 percent of global market cap, nearly four times larger than the second-place contender, China (SIFMA 2022). Outstanding US debt securities at the end of 2021, at $49.3 trillion, were more than double those of

16. Among recent studies are Linnemann and Schabert (2015), Engel (2016), Krishnamurthy and Lustig (2019), Valchev (2020), Jiang, Krishnamurthy, and Lustig (2021), Engel and Wu (forthcoming), and Lilley and others (2022). Relative "outside" bond supplies in global markets may change in the absence of monetary policy changes through balance sheet operations by government entities (including sterilized foreign exchange interventions) or through government fiscal imbalances.

17. Obstfeld (2015) documents the strong co-movement of global nominal long-term interest rates.

Figure 7. The US Dollar's Disproportionate Share in Global Assets and Transactions



Source: Adapted from Committee on the Global Financial System (2020) with some updated data.

the euro area or China.[18] Moreover, the US dollar's roles in world portfolios and transactions are unrivaled and go far beyond the United States' shares in world output or trade, as illustrated in figure 7.[19] By large margins, the dollar is the world's premier funding, reserve, invoice, anchor, and vehicle currency, an important reason for the outsized impact of US monetary and financial conditions on global activity. That impact is especially intense for EMDEs, which generally are more vulnerable to foreign financial shocks owing to shallower and less developed foreign exchange and capital markets, weaker financial regulatory frameworks, balance sheet weaknesses, and shorter track records of credible macro policies.[20]

18. Bank for International Settlements, "Debt Securities Statistics," https://www.bis.org/statistics/secstats.htm, accessed November 4, 2022.

19. An alternative source for recent data on the dollar's dominance is Bertaut, von Beschwitz, and Curcuru (2021). They analyze newer invoicing data assembled by Boz and others (2022) and find that the dollar's share in export invoicing is 96.3 percent in the Americas, 74.0 percent in the Asia-Pacific region, 23.1 percent in Europe, and 79.1 percent in the rest of the world. On the dollar's central and growing role in international bond markets, see Maggiori, Neiman, and Schreger (2020).

20. Gourinchas (2021) presents a comprehensive survey of the dollar's global roles. Models of the multiple network effects that underlie the dollar's unique position include Gopinath and Stein (2021), Chahrour and Valchev (2022), and Mukhin (2022). These types of models can also rationalize the dollar's exceptional liquidity or convenience yield. Bianchi, Bigio, and Engel (2021) model how the dollar's central role in international banking leads to a convenience premium and to dollar appreciation during global risk-off events. For theoretical models of US monetary policy transmission focusing on global safe dollar asset demand, see Canzoneri and others (2013), Jiang, Krishnamurthy, and Lustig (2020), and Kekre and Lenel (2021).

II. Emerging Markets and the Dollar

In this section we estimate the response to nominal US dollar appreciation for a sample of twenty-six EMDEs spanning multiple regions. The results indicate that dollar appreciation shocks are broadly contractionary, predicting prolonged downturns with the severity of the negative effects dependent on country characteristics.

II.A. Methodology and Initial Findings

Our core econometric exercise investigates how emerging market economies respond to changes in the nominal foreign exchange value of the US dollar. We proceed through a set of panel local projections (Jordà 2005):

$$y_{i,t+h} - y_{i,t-1} = \mu_{i,h} + \beta_h \Delta s_t + \gamma_h' \Delta z_t + \sum_{l=1}^{p} \delta_{h,l}' \Delta w_{i,t-l} + \varepsilon_{i,h,t}. \tag{1}$$

We unpack equation (1) term by term. The dependent variable is the cumulative change in country i's economic or financial variable y from quarter $t - 1$ to $t + h$, $h = 0, \ldots, H$. To understand the dollar's potentially pervasive influence on EMDEs more fully, we consider a wide range of economic indicators. To that end, we compile quarterly data for twenty-six EMDEs spanning the period from the late 1990s to 2019. While the makeup of our sample is largely dictated by data availability, it nonetheless covers about 90 percent of total 2021 EMDE GDP at market exchange rates and a time period that is reasonably uniform in terms of its high degree of global financial activity and integration. The data set includes information on national accounts, bilateral dollar exchange rates, related price indexes, terms of trade, domestic credit, equity prices, and interest rates. Here we report impulse responses for real GDP, investment, GDP deflator inflation, the bilateral exchange rate against the dollar, local currency equity prices, and the monetary policy interest rate. Online appendix A presents the full set of impulse response functions. Online appendix B provides a detailed report on the data sources for each country.

On the right-hand side of equation (1), a country- and horizon-specific intercept $\mu_{i,h}$ accounts for unobserved country heterogeneity as well as for linear trends in y. Our choice of shock variables and controls merits a detailed discussion. To measure shocks to the dollar exchange rate, Δs_t, we consider innovations to the trade-weighted dollar index against a basket of advanced economy (AE) currencies, obtained from the Federal Reserve H.10

release.[21] Typical emerging market economies will have little direct influence over the bilateral exchange rates among AE currency pairs, making the nominal AE dollar index plausibly external to EMDEs once appropriate controls have been imposed to account for common shocks to the aggregate of EMDEs that could feed back into the dollar's broad exchange rate against other AEs. The impulse response function of y is represented by the set of coefficients $\{\beta_h\}_{h=0}^{H}$.[22]

As we demonstrate further in section III and as a large body of literature has affirmed, dollar movements are highly responsive to various global and US-specific factors. Shifts in US monetary policy and financial conditions, as well as changes in investors' risk perceptions, can drive the dollar. At the same time, some of these factors are also endogenous and could respond to common shocks that hit the United States and foreign economies, including EMDEs. By including a vector of additional global controls Δz_t in equation (1), we get closer to a dollar shock component that is external to EMDE developments while allowing that other potential determinants of EMDE dynamics simultaneously have effects. Within z_t, we include US monetary policy as represented by the effective federal funds rate when the latter is positive and the Wu and Xia (2016) shadow rate during the zero lower bound period. As a way to control for US financial conditions, we adopt a factor-augmented approach by including in z_t the Federal Reserve Bank of Chicago's Adjusted National Financial Conditions Index (ANFCI). The index is constructed from a dynamic factor model of more than one hundred measures of financial activity in the United States and filters out the influence of overall economic activity and inflation.[23] In section III, we take a broader view and show that the dollar correlations reported in section I reflect the dollar's dependence on a range of shocks that potentially affect EMDEs.

Taken as a group, EMDEs are large enough that common EMDE shocks could potentially move the dollar exchange rate relative to other AEs. To reduce feedback from individual country outcomes to the dollar exchange

21. The currencies included in the Nominal Major Currencies U.S. Dollar Index (FRED ticker DTWEXM) are the euro, Japanese yen, Canadian dollar, UK pound sterling, Swiss franc, Australian dollar, and Swedish krona. We use quarter-end observations of the index with merchandise trade weights. We also check that our results are robust if we use quarterly averages of the index instead.

22. Using the terminology in Stock and Watson (2018), $\{\beta_h\}_{h=0}^{H}$ measures the cumulative impulse responses for first differences of the dependent variable.

23. For details on the ANFCI, see Brave and Kelley (2017). Our estimates are robust to alternative timing assumptions, in particular, if we control only for the lagged values of the US policy rate and financial conditions index.

rate through this channel, we control for aggregate economic activity in the EMDE bloc. Using a dynamic factor model like the one that underlies the ANFCI, we extract a common dynamic real GDP factor from an unbalanced quarterly panel of more than sixty EMDE countries. The intent of this additional global control, also included in Δz_t, is to capture EMDE business cycle fluctuations at a reasonably high frequency.[24]

Equation (1) also includes the vector of lagged controls $\Delta w_{i,t-l} \equiv (\Delta s_{t-l}, \Delta z_{t-l}, \Delta q_{i,t-l})'$, $l = 1, \ldots, p$, where the country-specific local controls $\Delta q_{i,t-l}$ comprise lags of $y_{i,t}$ as well as lags of additional country-specific economic indicators.[25] By lagging the local controls by one period, we implicitly make an ordering assumption: global controls and dollar shocks have instantaneous impacts on emerging economy variables, but the effects of EMDE economic and financial variables, including the policy responses to the dollar shock, themselves arrive with a lag.[26]

Our LP approach builds on several earlier contributions, all of which are informative but narrower than our analysis in various ways. Liu, Spiegel, and Tai (2017) explicitly apply a factor-augmented vector autoregressive (FAVAR) analysis to Korea, Japan, and China, but they display impulse responses based on a Cholesky ordering that precludes impact effects of dollar movements. Avdjiev, Bruno, and others (2019) include the nominal effective dollar in a panel vector autoregression (VAR) but examine a limited set of variables with no controls for global demand. Eguren Martin, Mukhopadhyay, and van Hombeeck (2017) and Hofmann and Park (2020) come closer to our suggested method but examine a limited range of response variables. Eguren Martin, Mukhopadhyay, and van Hombeeck (2017) focus on growth outcomes only, while Hofmann and Park (2020) are largely concerned with the dollar's connection with expected distributions of future investment and exports. The closest precursor to our approach is Shousha (2022), who investigates the EMDE response to dollar shocks through a

24. Online appendix B provides an overview of the model and estimation method. Figure A7 plots our estimated dynamic emerging market demand factor.

25. Specifically, we include lagged quarterly changes in real GDP, the bilateral exchange rate against the US dollar, and the policy interest rate. As these controls have long data series often extending back to the 1980s, we ensure that our LP procedure utilizes as much data as possible, while avoiding over-parameterizing the model by including too many controls. Our estimate corresponds to the "lag-augmented" LP estimator of a VAR(p) model for the data $(y, q, s, z)'$ (Montiel Olea and Plagborg-Møller 2021). The lag-augmented approach allows us to compute Eicker-Huber-White standard errors for robust inference over potentially nonstationary data. We choose a conservative VAR lag by setting p equal to four quarters.

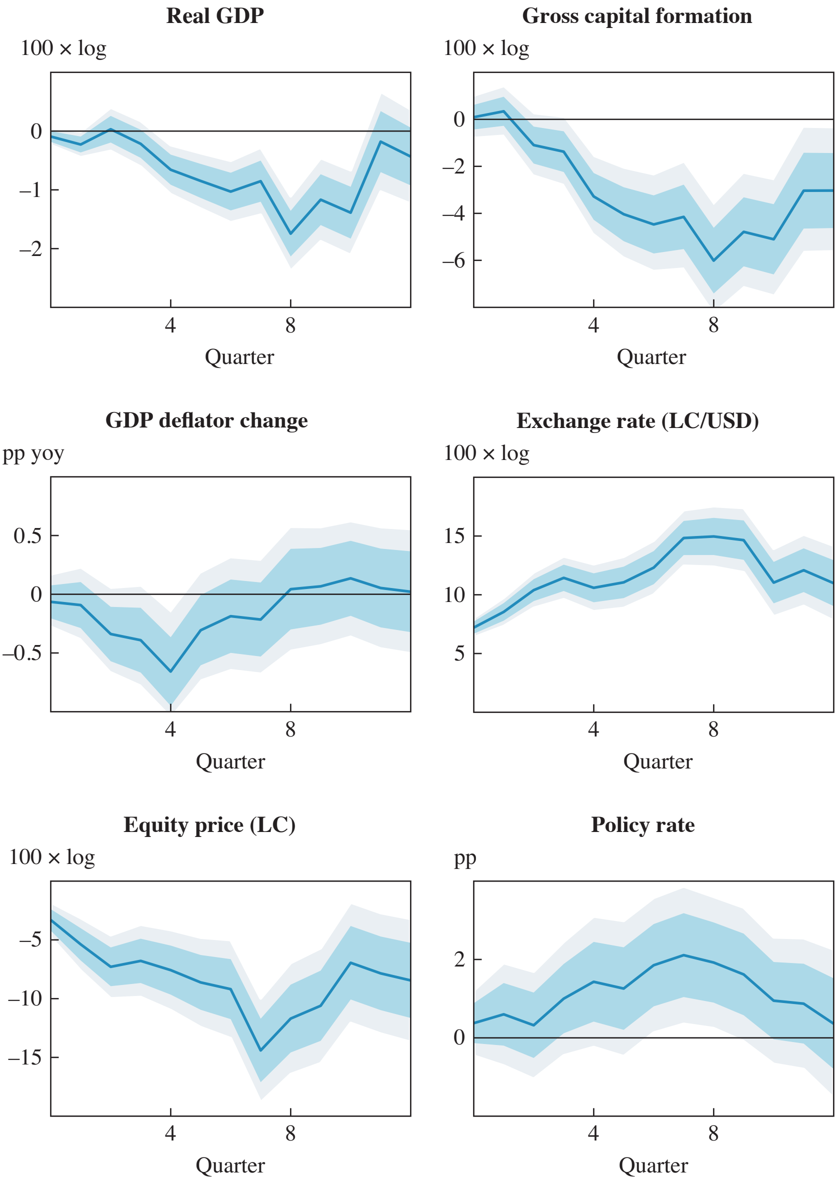
26. Plagborg-Møller and Wolf (2021) discuss the implementation of structural vector autoregressive (SVAR) restrictions in local projections.

VAR model. While our findings in this section are broadly similar and complementary, we push our analysis further in several ways. We use a flexible yet robust LP approach on a larger country sample and examine a wider range of EMDE outcome variables. By focusing on the dollar's exchange rate against AEs only and adding factor-augmented controls, we obtain a sharper identification of dollar shocks that are external to developments in EMDEs. Like Shousha (2022), we also consider potential country-level heterogeneity in the transmission of dollar shocks. As will be clear in section II.B, our state-dependent LP estimation is more flexible in explicitly accommodating time variation in policy regimes and balance sheet exposures.

Figure 8 shows the average response to a 10 percent dollar appreciation in our EMDE sample. We report impulse response functions as well as 68 percent and 90 percent confidence bands. In response to the dollar shock, real GDP falls, reaching a trough of about −1.5 percent relative to trend after about eight quarters. In line with this output response, investment also falls. Year-over-year inflation in the GDP deflator falls over four quarters before starting to recover. The domestic currency depreciates immediately against the dollar. This bilateral depreciation continues subsequently, reversing partially only after output bottoms out. In online appendix A.1, we show that in line with a contraction in global trade, export and import prices both decline. However, export prices lose more ground than import prices, so the terms of trade deteriorate and reinforce other contractionary forces on spending. For indicators of financial market responses, the central bank policy rate is estimated to rise marginally on impact and subsequently it rises further for about two years. While this estimate is not statistically significant until several quarters have passed, there are additional financial repercussions through a sharp fall in equity prices, as well as a rise in the emerging markets bond index (EMBI) spread on sovereign dollar borrowing and a decline in nominal domestic credit (both shown in online appendix A.1). These all contribute to the overall contractionary impact of the dollar shock.[27]

27. Adopting the Gorodnichenko and Lee (2020) methodology for variance decompositions in LPs, we find an important role for dollar shocks in explaining the dynamics of macro aggregates in our sample of emerging market economies. For consumption, exports, and aggregate output, the shares explained by dollar shocks reach 25 to 30 percent after two quarters. On the financial side, dollar appreciation explains around 20 percent of equity price variance after eight quarters.

Figure 8. Impulse Response: 10 Percent Appreciation of Advanced Economies' Dollar Index



Source: Authors' calculations.

Note: The impulse response functions of EMDE economic and financial variables to a 10 percent appreciation of the dollar exchange rate against a basket of advanced economy currencies, based on the local projection, equation (1). For regressions involving the GDP deflator, country-quarter observations with a year-over-year change greater than 50 percent are dropped. Equity prices are local currency stock market indexes. Heteroskedasticity-robust 90 percent and 68 percent confidence bands are reported.

II.B. Dollar Shocks and Country Heterogeneity

Following a series of studies starting with Ramey and Zubairy (2018), we extend our LP framework to allow the impact of dollar shocks to differ based on predetermined characteristics or "states" of EMDEs. Formally, we estimate the following panel LP with state dependence:

$$(2) \qquad y_{j,t+h} - y_{j,t-1} = I_{j,t-1} \times \left[\mu_{A,j,h} + \beta_{A,h} \Delta s_t + \gamma'_{A,h} \Delta z_t + \sum_{l=1}^{p} \delta'_{A,h,l} \Delta w_{j,t-l} \right] + \left(1 - I_{j,t-1}\right) \times \left[\begin{array}{c} \mu_{B,j,h} + \beta_{B,h} \Delta s_t + \gamma'_{B,h} \Delta z_t \\ + \sum_{l=1}^{p} \delta'_{B,h,l} \Delta w_{j,t-l} \end{array} \right] + \varepsilon_{j,h,t}.$$

The indicator function $I_{j,t-1}$ takes the value 1 if country j's economy is in state A on date $t - 1$ (that is, prior to the shock realization Δs_t) and 0 if it is in state B.[28] The slope coefficients associated with $I_{j,t-1} \cdot \Delta s_t$ in state A, $\{\beta_{A,h}\}_{h=0}^{H}$, can be interpreted as the impulse response function conditional on the economy being in that state and similarly for $\{\beta_{B,h}\}_{h=0}^{H}$ and state B.

Ex ante policy regimes and external balance sheet exposure to dollar movements define states of the economy prominent in policy discussions of EMDEs' vulnerability to dollar shocks. We consider three dimensions of country heterogeneity: flexibility of the exchange rate, whether the central bank is an inflation targeter (as a proxy for monetary policy credibility), and the degree of dollar denomination of liabilities to foreigners.

The findings in this section should be interpreted with caution because countries are not allocated randomly among policy or financial regimes. Perhaps countries with different degrees of foreign dollar liability exposure also differ in other respects. For example, if countries with more dollar exposure also trade more with the United States, their trade might be affected more strongly by dollar shocks for reasons unconnected with financial structure. Another potential bias comes from the endogeneity of policy regimes. Some countries might choose their exchange rate regime with an eye toward minimizing impacts from the external shocks that they face. In that case, we might underestimate the contrasts between more and less flexible exchange rate regimes. Countries that adopt inflation targeting

28. In the international macro literature, Ben Zeev (2019) uses a state-dependent LP framework to study the interaction between international credit supply shocks and the exchange rate regime. Recent work by Gonçalves and others (2022) establishes the validity of the state-dependent LP approach, in particular if the state indicators depend only on lagged endogenous variables. As our discussion suggests, our choices of states are likely to satisfy that requirement.

might simply be those endowed with a range of other institutional features that would enhance macro stability even without a formal inflation target.

EXCHANGE RATE FLEXIBILITY Countries with more exchange rate flexibility have an extra degree of freedom to respond to global shocks. The exchange rate itself is to some extent a two-edged weapon: depreciation in the face of a negative external impulse can raise aggregate demand for domestic goods through the net export channel and also raise trade-oriented firms' demand for labor and new capital, but it may damage balance sheets with contractionary effects.[29] However, a flexible exchange rate frees the central bank to move policy interest rates independently of foreign rates so as to stabilize the economy, and it removes the need for measures to defend a pegged exchange rate against speculative attacks.[30]

Rey (2013) argued that the global financial cycle to some degree renders the choice of exchange rate regime for EMDEs moot, since even a floating rate cannot repel financial shocks coming from advanced financial markets. However, a number of empirical studies suggest that even for EMDEs, more flexible regimes mitigate the adverse effects of various global shocks like the dollar shock responses we documented above, even if they do not fully offset them. We will add support to that view.[31]

We define countries as having exchange rate pegs according to Ilzetzki, Reinhart, and Rogoff's (2019) classification. In our application, we consider an exchange rate as pegged when it is either a fixed peg or a crawling peg with narrow bands in the final month of a quarter.[32] Other countries, either

29. Even when exports are invoiced in dollars, so that domestic currency depreciation does not immediately lower export prices for foreigners and thereby spur higher foreign demand, exporter profits rise, encouraging hiring, consumption, and investment.

30. Kalemli-Özcan (2019) makes a related argument. She shows that a contractionary US monetary shock raises the required excess return on EMDE bonds, a contractionary effect. Under a flexible exchange rate, this risk premium increase is achieved in part through an immediate currency depreciation. Under a pegged exchange rate, however, a sharper domestic monetary contraction would be needed to achieve the same risk premium rise, with even more damage to the economy.

31. For example, Obstfeld, Ostry, and Qureshi (2019) consider shocks to the CBOE S&P 100 Volatility Index (VXO, the precursor of the VIX); Loipersberger and Matschke (2022) consider shocks to the CBOE Volatility Index (VIX); Ben Zeev (2019) considers shocks to the EBP; and Degasperi, Hong, and Ricco (2021) consider shocks to US monetary policy. Gourinchas (2018) estimates a model of the Chilean economy incorporating potential expansionary and contractionary channels of peso depreciation and concludes that, on balance, exchange rate flexibility supports the central bank's stabilization efforts.

32. That is, our pegs have coarse classification codes 1 and 2 (Ilzetzki, Reinhart, and Rogoff 2019). Loipersberger and Matschke (2022) also adopt this definition of a pegged rate. Emerging European economies whose currencies are anchored or pegged to the euro are regarded as having a flexible exchange rate against the dollar. Observations designated as a "free-falling" or "dual-market" exchange rate regime are dropped from our analysis.

freely floating their currencies or having relatively more flexible currency managements, are labeled as floaters.

Figure 9 shows the response to a 10 percent dollar appreciation according to the flexibility of the exchange rate regime. GDP and investment fall more sharply for countries with exchange rate pegs, consistent with the idea that exchange rate flexibility helps buffer dollar shocks. There is a significant fall in the GDP deflator for pegs. The stock market also drops more sharply in pegs. Countries with exchange rate pegs are more likely to raise their policy interest rates in the short run and over time to maintain their exchange rates, possibly contributing to the deflationary force of the dollar shock. In contrast, countries with floats do not tighten monetary policy in response to contractionary dollar shocks.[33] Countries with pegs display a smaller currency depreciation over the first year or so (as one would expect) and bigger falls in export prices and the terms of trade (see online appendix A.1).[34]

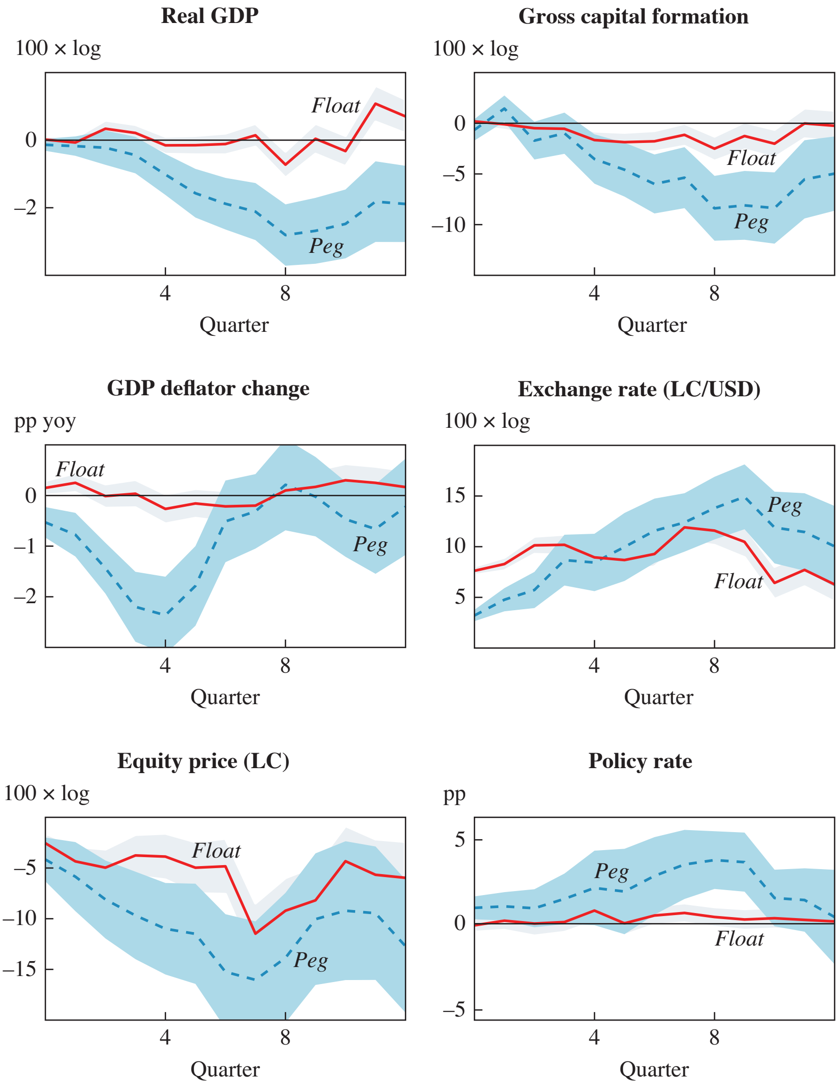
The general picture that emerges is one in which countries with more exchange rate flexibility do better in coping with the external shock of dollar appreciation.

MONETARY POLICY CREDIBILITY Flexible exchange rates can also promote macroeconomic stability by enhancing monetary autonomy and thereby allowing the adoption of a credible inflation-targeting regime. Moreover, when monetary policy is credible, a central bank can allow exchange rate fluctuations to buffer the economy against foreign shocks with less worry about de-anchoring inflation expectations or rapid exchange rate pass-through to domestic prices (Bems and others 2021). Thus, we expect that inflation-targeting EMDEs may fare better in the face of dollar shocks from abroad. In defining the inflation-targeting state indicator for our estimates, we adopt

33. De Leo, Gopinath, and Kalemli-Özcan (2022) document that EMDE central banks with more flexible exchange rates cut their policy interest rates in response to instrumented US monetary policy shocks (Gertler and Karadi 2015) and argue that EMDE monetary responses have therefore tended to be countercyclical, consistent with the findings on sudden capital inflow stops in Eichengreen and Gupta (2018). However, our notion of dollar shocks is broader than that of Gertler and Karadi (2015), which accounts for only a small share of dollar variability, or sudden stops.

34. As the online appendix also shows, domestic credit rises initially in countries with pegs, which could reflect a countercyclical policy attempt under the constraint of a peg. Remember that our definition of "peg" includes crawling bands, which therefore may respond to shocks over time. Export prices would fall less for floaters if, as the data in Boz and others (2022) suggest is true for many EMDEs, exports are invoiced in dollars, so that a depreciation of the domestic currency against the dollar pulls their domestic-currency prices up relative to the case of pegs.

Figure 9. Impulse Response: 10 Percent Appreciation of Advanced Economies' Dollar Index, by FX Regime



Source: Authors' calculations.

Note: The impulse responses of EMDE economic and financial variables to a 10 percent dollar appreciation against a basket of advanced economy currencies, conditional on the exchange rate regime. Estimates are derived from the state-dependent local projection, equation (2). The state indicator I_{t-1} is defined based on the Ilzetzki, Reinhart, and Rogoff (2019) (IRR) exchange rate regime one quarter prior to the current quarter t. A country is considered to have a floating exchange rate ($I_t = 1$) if it is assigned an IRR coarse regime code of 3 or 4 in quarter t. Countries with a pegged exchange rate have an IRR coarse regime code of 1 or 2. The figure plots 68 percent robust standard error bands. For regressions involving the GDP deflator, country-quarter observations with year-over-year change greater than 50 percent are dropped. Equity prices are local currency stock market indexes.

the classification of Ha, Kose, and Ohnsorge (2021), which is based on the IMF's *Annual Report on Exchange Arrangements and Exchange Restrictions* database.[35]

Figure 10 shows how the impulse responses differ depending on the monetary regime. For macro aggregates such as real GDP and investment, the results are broadly similar to the pegged/float comparison in figure 9. In non-targeters, however, there is more deflation over time, the bilateral currency depreciation against the dollar is greater over time, and the stock market slump is deeper. Non-targeters raise their policy interest rates, which is consistent with a stronger deflationary response. In online appendix A.1, we show that the terms of trade evolve similarly for the two groups. In addition, non-targeters see a bigger contraction in domestic credit and soon see rises in their EMBI spreads.

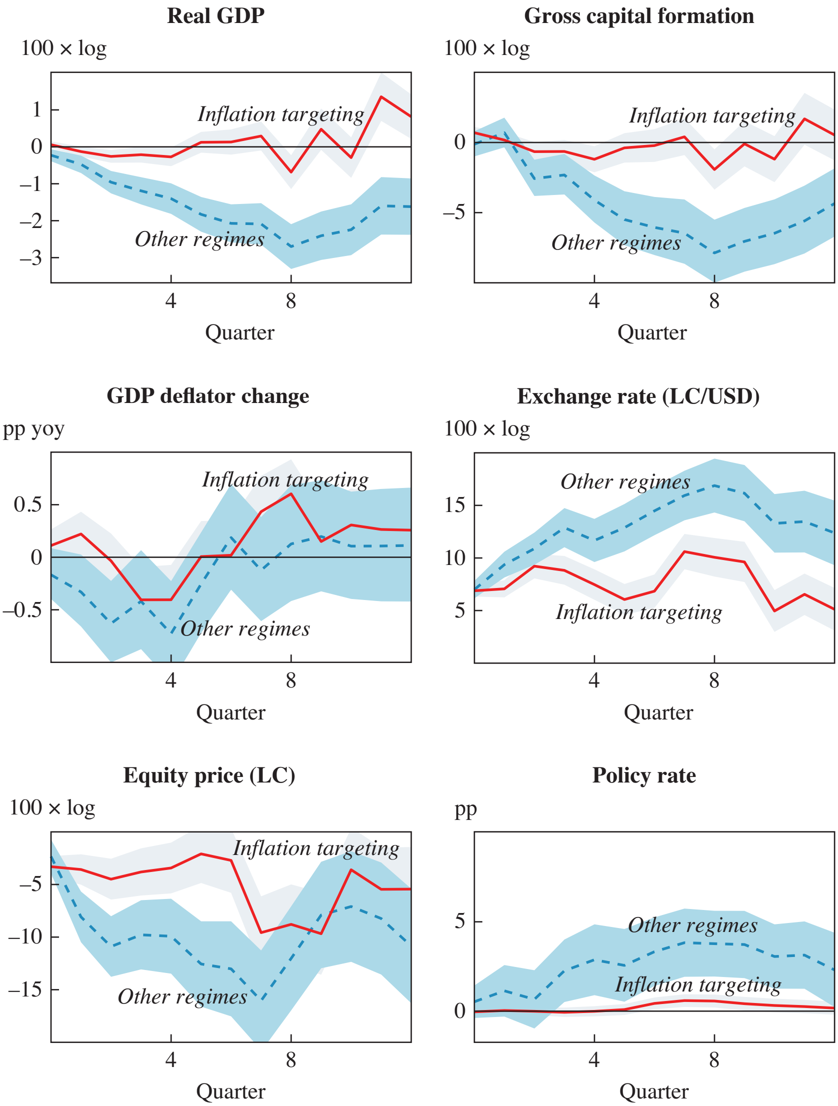
DOLLAR LIABILITIES Finally, EMDEs with large dollar-denominated liabilities are potentially vulnerable to unexpected domestic currency depreciation against the dollar that increases real debt burdens. Less dollarization of external liabilities should mitigate the procyclical effects of dollar movements on domestic balance sheets and financial conditions (especially when the exchange rate is more flexible).

We use Bénétrix and others' (2019) estimates of the currency composition of external positions to gauge the role of external balance sheet exposure to adverse dollar appreciation. The indicator $I_{j,t-1}$ takes the value 1 if during year $t - 1$, country j's dollar-denominated portfolio liabilities as a share of GDP exceed the median over all country-time observations in our twenty-six-country sample.

Figure 11 shows that when the dollar appreciates, countries with higher external dollar exposure suffer bigger declines in GDP after about four quarters. Incongruously, investment is predicted to rise initially and remain higher in high-exposure countries. High-exposure countries eventually experience greater depreciation against the dollar and see steeper equity-price declines and bigger hikes in policy rates. Online appendix A.1 reports that high-exposure countries suffer a significantly larger adverse terms of trade change, and also display slower domestic credit growth after about four quarters. Finally, high-exposure countries experience persistently higher EMBI sovereign spreads.

35. Our data on monetary regimes and dollar liabilities (see the next subsection) run until the end of 2017.

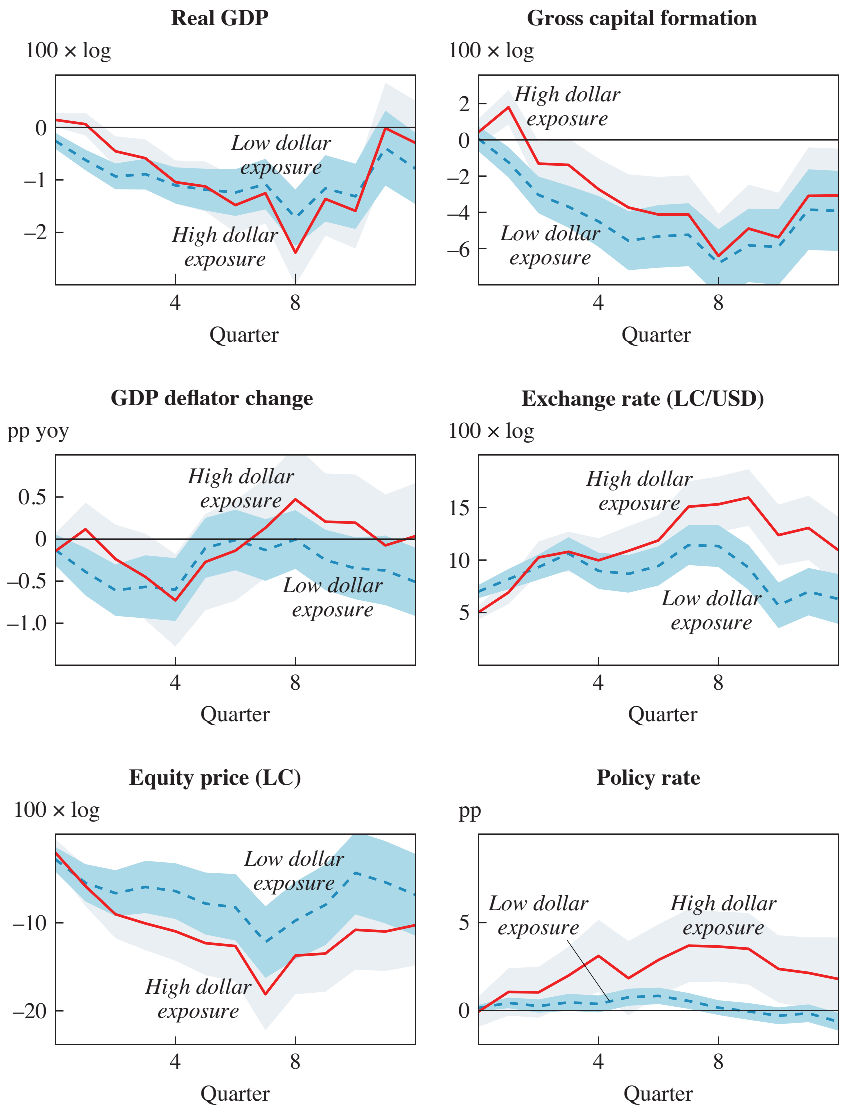
Figure 10. Impulse Response: 10 Percent Appreciation of Advanced Economies' Dollar Index, by Monetary Regime



Source: Authors' calculations.

Note: The impulse responses of EMDE economic and financial variables to a 10 percent dollar appreciation against a basket of advanced economy currencies, conditional on the monetary policy regime. Estimates are derived from the state-dependent local projection, equation (2). The state indicator I_{t-1} is defined based on the classification of Ha, Kose, and Ohnsorge (2021). A country is in state $I_{t-1} = 1$ only if it practices inflation targeting in the previous year. The figure plots 68 percent robust standard error bands. For regressions involving the GDP deflator, country-quarter observations with year-over-year change greater than 50 percent are dropped. Equity prices are local currency stock market indexes.

Figure 11. Impulse Response: 10 Percent Appreciation of Advanced Economies' Dollar Index, by Dollar Liability to GDP



Source: Authors' calculations.

Note: The impulse responses of EMDE economic and financial variables to a 10 percent dollar appreciation against a basket of advanced economy currencies, conditional on the degree of balance sheet exposure to the dollar. Estimates are derived from the local projection, equation (2). The state indicator I_{t-1} is based on the cross-border currency exposure data set of Bénétrix and others (2019). A country is in state $I_{t-1} = 1$ if its external dollar liabilities as a share of GDP in the previous year exceed the median of all country-quarter observations. The figure plots 68 percent robust standard error bands. For regressions involving the GDP deflator, country-quarter observations with year-over-year change greater than 50 percent are dropped. Equity prices are local currency stock market indexes.

SUMMARY More exchange rate flexibility, an inflation-targeting monetary framework, and lower dollar liabilities to foreigners all generally strengthen an emerging economy's defenses against a dollar appreciation shock. Other features of an economy can be important as well. Shousha's (2022) findings suggest that lower dollar invoicing of exports and greater integration into global value chains enhance macro stability. He reports similar results to ours concerning exchange rate flexibility and monetary policy credibility.

We have also examined the role of openness to cross-border financial flows, asking whether restrictions on capital flows enhance resilience to external dollar shocks. Using the Chinn and Ito (2006) de jure measure of financial openness, we examined the response to a dollar shock in EMDEs with relatively open and closed financial accounts.[36] Capital flow restrictions appear to make little difference for the effects on real variables or the exchange rate, but countries with higher openness experience bigger rises in short-term interest rates and EMBI spreads, along with a significantly bigger fall in domestic credit. This evidence needs to be interpreted with caution, but it suggests that the stabilization benefits from capital controls may be smaller than those from exchange rate flexibility, credible monetary policy, and avoidance of external dollar liabilities.[37]

III. Financial Determinants of the Dollar Exchange Rate

Movements in the US dollar's effective nominal exchange rate against advanced economies clearly have an impact on EMDEs. The dollar's influence appears stronger in countries with more rigid exchange rate regimes, less credible monetary frameworks, and more foreign currency external debt. Those findings give a partial insight into the correlations of EMDE

36. We classify a country as relatively open if its normalized Chinn-Ito score, ranging from 0 (most closed) to 1 (most open), exceeds 0.5. For example, Indonesian measures pushed the country from a score of 0.70 in 2010 to 0.42 in 2011; Brazil moved from 0.48 in 2005 to 0.54 during 2006–2009 and as far down as 0.16 by 2015.

37. Even for China, which maintains a relatively high level of capital flow controls but manages its exchange rate, the annual correlation between real output growth and nominal dollar appreciation is –0.50 over 1999–2021. Over the same period, the correlation of China's growth rate with that of EMDEs other than China (based on the IMF's PPP-weighted growth measure) is about –0.8. A more granular treatment of controls would differentiate between inflow and outflow controls. Consistent with our findings, Klein and Shambaugh (2015) find that capital controls, unless extensive, do little to enhance the efficacy of monetary policy. Loipersberger and Matschke (2022) conclude that capital controls can yield stabilization benefits for EMDEs with pegged, but not floating, exchange rate regimes.

activity with the dollar reported in section I. Insight into the channels of dollar influence comes from identifying shocks that drive the broad nominal dollar.

III.A. Modeling the Dollar's Exchange Rate against Advanced Economies

To model the dollar's exchange rate against advanced economies, we follow Engel and Wu (forthcoming) and start with a modified interest parity relationship.[38] Let s denote the log dollar exchange rate, defined as the foreign currency price of the dollar, so that a rise in s is an appreciation of the dollar. Let i_t^L denote the interest rate per period on a short-term market dollar instrument,for example, the London Interbank Offered Rate (LIBOR), and i_t^L* the interest rate per period on a comparable foreign currency instrument. The classic UIP condition, based on risk neutrality, full arbitrage, and rational expectations, is written:

$$i_t^{L*} - \left(i_t^L + \mathbb{E}_t s_{t+1} - s_t\right) = 0. \tag{3}$$

There is extensive evidence against this simple form of interest parity. We modify it by introducing two additional factors. Let ρ_t denote an equilibrium excess return on the trade in which one borrows dollars and invests in interest-bearing foreign currency assets. As noted above, the excess return may result simply from optimization under risk aversion, in which case it might reflect the covariance of the dollar's value with a stochastic discount factor, but it could alternatively be a required net return on investment determined by incentive constraints (Gabaix and Maggiori 2015) or a combination of these elements (Gabaix and Maggiori 2015; Itskhoki and Mukhin 2021). Also in play might be heterogeneous expectations that diverge from well-informed rational expectations. We denote by $\lambda_t^{\$}$ an additional liquidity or convenience yield on the dollar instrument (relative to foreign currency instruments) owing to the dollar's unique global role. The modified UIP condition would then read:

$$i_t^{L*} - \left(i_t^L + \mathbb{E}_t s_{t+1} - s_t\right) = \rho_t + \lambda_t^{\$}.$$

38. Exchange rate models of the 1970s, such as Dornbusch (1976), also started from interest parity but, in monetarist fashion, emphasized relative money supplies as an ultimate driver of relative interest rates and thereby of exchange rates. More recent models recognize interest rates as instruments of monetary policy and therefore as direct drivers of exchange rates. We take that approach here.

This equation can be solved forward to express the exchange rate's current level in terms of expected future interest rate differences, excess returns, dollar liquidity shocks, and a terminal exchange rate:

$$(4) \qquad s_t = \sum_{s=0}^{k-1} \mathbb{E}_t \left(i_{t+s}^L - i_{t+s}^{L*} \right) + \sum_{t=0}^{k-1} \mathbb{E}_t \left(\rho_{t+s} - \lambda_{t+s}^{\$} \right) + \mathbb{E}_t \left(s_{t+k} \right).$$

A skeptical view of equation (4) would be that the composite term $\rho_t + \lambda_t^{\$}$ is "dark matter" that tautologically gives an interest parity–based theory of the exchange rate empirical validity. The theory acquires content from measurable correlates of ρ_t and $\lambda_t^{\$}$ that can be justified by empirically persuasive models. In general, it is challenging to identify effects of the two shocks individually, as they surely are driven by common factors. For example, a rise in global safe asset demand due to higher risk aversion could be associated with a simultaneous tightening of balance sheet constraints and rise in the marginal convenience value of dollars, leading to positive co-movement in ρ_t and $\lambda_t^{\$}$.[39]

Further insights into the determinants of exchange rates come from considering the liquidity advantages of safer government-issued bonds compared with privately issued market instruments. We denote by $i_t(i_t^*)$ the US (foreign) short-term central government bond yield. If $i_t^L - i_t(i_t^{L*} - i_t^*)$ is taken to measure the marginal liquidity yield on the US Treasury (foreign government) liability, then we may take:

$$\gamma_t \equiv i_t^L - i_t - \left(i_t^{L*} - i_t^* \right)$$

as a measure of relative Treasury liquidity, as suggested by Engel and Wu (forthcoming). Importantly, γ_t differences out the pure relative liquidity value of dollar denomination captured by $\lambda_t^{\$}$. The last definition, together with equation (4), allows us to express the exchange rate in terms of relative government bond yields as:

$$(5) \qquad s_t = \sum_{s=0}^{k-1} \mathbb{E}_t \left(i_{t+s} - i_{t+s}^* \right) + \sum_{t=0}^{k-1} \mathbb{E}_t \left(\rho_{t+s} + \lambda_{t+s}^{\$} + \gamma_{t+s} \right) + \mathbb{E}_t \left(s_{t+k} \right).$$

Equation (5) will provide one basis for our empirical study of correlates of the dollar's exchange rate, but there are two other versions of the exchange

39. As Krishnamurthy and Lustig (2019) put it, convenience yields are relevant even when intermediaries are unconstrained, but "innovations to the convenience yield are certainly correlated with shocks to the financial sector" (456).

rate equation that provide complementary perspectives. Let $i_t^{(k)}(i_t^{(k)*})$ be the k-period long-term Treasury (foreign government bond) zero coupon yield. According to a standard approximation, $i_t^{(k)}$ is related to the path of expected future short rates by:

$$i_t^{(k)} = \frac{1}{k}\sum_{s=0}^{k-1}\mathbb{E}_t\left(i_{t+s}\right) + \tau_t^{(k)},$$

where $\tau_t^{(k)}$ is the term premium on a k-period US government bond. A corresponding equation involving the foreign term premium $\tau_t^{(k)*}$ holds for the foreign government bond. Using the term structure relationships, we express equation (5) as:

$$(6) \qquad s_t = k\left(i_t^{(k)} - i_t^{(k)*}\right) - k\left(\tau_t^{(k)} - \tau_t^{(k)*}\right) + \sum\nolimits_{t=0}^{k-1}\mathbb{E}_t\left(\rho_{t+s} + \lambda_{t+s}^{\$} + \gamma_{t+s}\right) + \mathbb{E}_t\left(s_{t+k}\right).$$

A final relationship comes from explicitly considering cross-currency arbitrage in long-term bonds. Denoting the annualized excess return and liquidity factors on k-period long-term government bonds by $\rho_{t+s}^{(k)}$, $\lambda_{t+s}^{(k)\$}$, and $\gamma_{t+s}^{(k)}$, we translate the longer-term interest parity relationship into an expression for the current spot exchange rate:

$$(7) \qquad s_t = k\left(i_t^{(k)} - i_t^{(k)*} + \rho_t^{(k)} + \lambda_t^{(k)\$} + \gamma_t^{(k)}\right) + \mathbb{E}_t\left(s_{t+k}\right).$$

Equations (5), (6), and (7) lead to different (but related) estimation specifications, given empirical stand-ins for the deviations from strict UIP.[40] For example, let Δ denote a first difference (which in practice will be a three-month or one-year first difference resulting in overlapping monthly observations).[41] Equation (5) suggests the specification:

$$(8) \qquad \Delta s_t = \alpha + \beta_1\Delta\left(i_t - i_t^*\right) + \beta_2\Delta\rho_t + \beta_3\Delta\lambda_t^{\$} + \beta_4\Delta\gamma_t + X_{t-1}\delta + \varepsilon_t,$$

where X_{t-1} contains lagged (by three or twelve months) levels of the included variables, as well as lagged variables useful in predicting the

40. We will not attempt to explore the constraint implied by equations (6) and (7), that $\rho_t^{(k)} + \lambda_t^{(k)\$} + \gamma_t^{(k)} = \frac{1}{k}\sum_{s=0}^{k-1}\mathbb{E}_t(\rho_{t+s} + \lambda_{t+s}^{\$} + \gamma_{t+s}) - (\tau_t^{(k)} - \tau_t^{(k)*})$.

41. This practice is also adopted by Hansen and Hodrick (1980), Greenwood and others (2020), and Dahlquist and Söderlind (2022), among others. We further ensure consistency with theory by matching the tenors of interest rates and currency bases wherever possible.

included first differences. The error term ε_t contains the expectations innovation $\mathbb{E}_t s_{t+k} - \mathbb{E}_{t-1} s_{t+k}$, likely to be small for large k, as well as any omitted date t shocks explaining revisions to the right-hand side of equation (5). While equation (8) therefore cannot be viewed as a structural relationship, it still yields useful information on the empirical correlates of dollar movements. One variable we include in the matrix X_{t-1} is the lagged log real exchange rate, which Eichenbaum, Johannsen, and Rebelo (2021) find to be a powerful predictor of future changes in the nominal exchange rate.[42] Using equation (6) and an approximation suggested by Du, Pflueger, and Schreger (2020), we derive an alternative regression equation:[43]

$$(9) \qquad \Delta s_t = \alpha + \beta_1 k \Delta\left(i_t^{(k)} - i_t^{(k)*}\right) + \beta_2 k\left(\tau_t^{(k)} - \tau_t^{(k)*}\right) + \beta_3 \Delta\rho_t + \beta_4 \Delta\lambda_t^{\$} + \beta_5 \Delta\gamma_t + X_{t-1}\delta + \varepsilon_t,$$

where we replace the short-term government yield differential in the lagged control X_{t-1} by the long-term government yield differential and the term premium differential.

Finally, equation (7) suggests the formulation:

$$(10) \qquad \Delta s_t = \alpha + \beta_1 k \Delta\left(i_t^{(k)} - i_t^{(k)*}\right) + \beta_2 k \Delta\rho_t^{(k)} + \beta_3 k \Delta\lambda_t^{(k)\$} + \beta_4 k \Delta\gamma_t^{(k)} + X_{t-1}\delta + \varepsilon_t.$$

Empirical exchange rate studies have generally focused on short-term interest rates as in equation (8), but large-scale central bank purchases of long-term bonds since the global financial crisis have rekindled interest in the role of long-term rates, as captured in equations (9) and (10). Models by Greenwood and others (2020) and Gourinchas, Ray, and Vayanos (2022), for example, argue that increases in a country's supply of long-term government bonds will push long-term interest rates up and appreciate its currency,

42. We take no stand on whether the nominal exchange rate log level is a stationary or nonstationary random variable. Jiang, Krishnamurthy, and Lustig (2021) assume it is stationary, whereas Engel and Wu (forthcoming) assume it is not, and both agree that the real exchange rate is stationary, if highly persistent. Itskhoki (2021), on the other hand, argues that real exchange rates are nonstationary. Mindful that our exchange rate equations are not structural, we would nonetheless assume that revisions to nominal exchange rate expectations far in the future have minimal correlation with current financial variables, for which stationarity is sufficient but not necessary.

43. In particular, we approximate $i_t^{(k+1)}$ by $i_t^{(k)}$ and $\tau_t^{(k+1)}$ by $\tau_t^{(k)}$ at quarterly and yearly horizons. Intuitively, the yield curve at long tenors is relatively flat.

whereas central bank purchases (which withdraw bonds from the market) will result in lower long-term rates and depreciation. In contrast, the analyses in Krishnamurthy and Lustig (2019) and Jiang, Krishnamurthy, and Lustig (2021) suggest that increases in US long-term bond supplies could push the currency down by reducing the marginal convenience yields represented by $\gamma_t^{(k)}$ and $\lambda_t^{(k)\$}$ in equation (7).

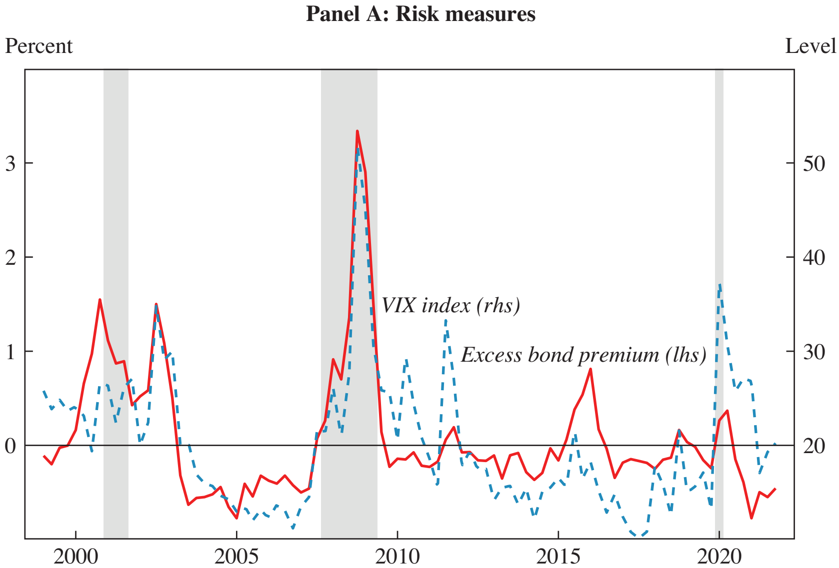
We will not try to resolve the general equilibrium effects of long-term bond purchases here but will simply document the correlations of the dollar exchange rate with proxies for the main determining factors. Chief among these are long-term interest rates themselves, which we derive from estimated zero coupon yield curves from Bloomberg. We also use the zero coupon yield curves to extract term premia, based on Adrian, Crump, and Moench's (2013) term structure model.[44] Figure A8 in the online appendix plots our estimated term premium series for each country and compares them with other term premium estimates in the literature.

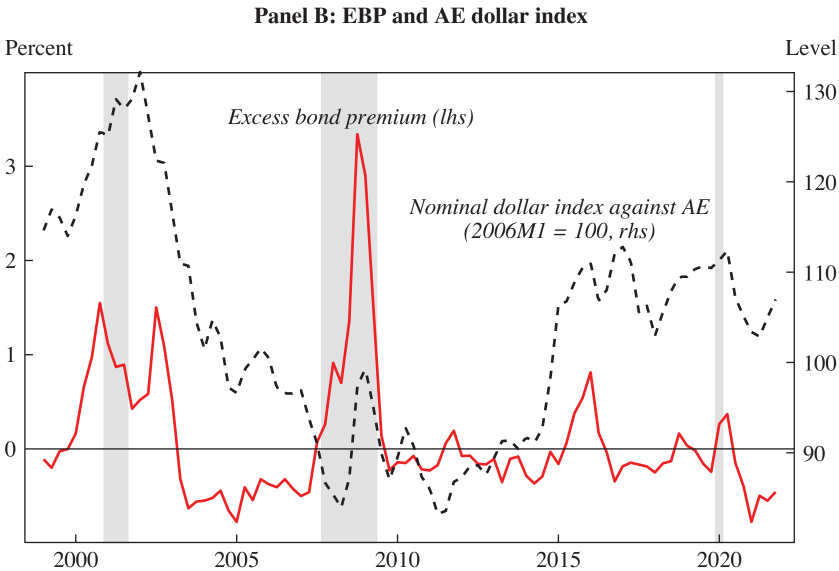
In estimating equations (8)–(10), we use two proxy variables to capture potential variation in the excess return terms, the Chicago Board Options Exchange Volatility Index (VIX) and the EBP of Gilchrist and Zakrajšek (2012). The VIX appears in many studies to capture generalized shifts in global risk aversion.[45] As Gilchrist and Zakrajšek (2012) explain, the EBP is built up from individual US corporation bond spreads, adjusted to remove estimates of firm-specific default risk and thus reflecting risk appetite or market sentiment rather than expected cash flows. Lilley and others (2022) find roles for related variables in explaining the variation of the dollar exchange rate after the global financial crisis, and all of them arguably are indicators of financial stresses that could have an impact on required excess returns, as well as liquidity convenience yields. Figure 12

44. Greenwood and others (2020) argue that foreign assets and long-term US government bonds are portfolio substitutes because they are similarly exposed to US short-term interest rate risk, which generally will move foreign exchange asset values and US bond prices in the same direction. Thus, when the supply of US long-term bonds rises, investors will want to sell foreign long-term assets as they rebalance their portfolios, making the dollar appreciate. The "original sin redux" argument of Carstens and Shin (2019) suggests there would be especially high substitutability between US long-term Treasuries and long-term sovereign EMDE bonds. In contrast, short-maturity US bonds and foreign assets are more complementary in portfolios owing to the diversification motive. One challenge in determining empirically the exchange rate effects of bond operations like quantitative easing (QE) is that they also can signal central bank targets for the price level path, with effects on future expectations of inflation and nominal interest rates.

45. Examples include Forbes and Warnock (2012), Rey (2013), Obstfeld, Ostry, and Qureshi (2019), Kalemli-Özcan (2019), Kalemli-Özcan and Varela (2021), and Loipersberger and Matschke (2022).

Figure 12. Key Proxy Drivers of Excess Returns: Quarterly Averages





Sources: Gilchrist and Zakrajšek (2012); FRED; Federal Reserve H.10 release.

Note: Panel A plots the evolution of the Gilchrist and Zakrajšek (2012) EBP (left-hand *y* axis, extracted from US nonfinancial firms' borrowing spreads) and the CBOE VIX (right-hand *y* axis). Panel B plots the Federal Reserve H.10 nominal dollar index against advanced economy currencies along with EBP. Shaded areas correspond to US recession episodes as dated by the National Bureau of Economic Research (FRED ticker USRECM).

plots the VIX and EBP measures and compares them with the broad dollar index.

For γ_t we use alternative measures of low- or no-risk private sector borrowing spreads over government bond rates. At the three-month horizon we use the difference between the TED spread (of LIBOR over the US Treasury bill rate) and its foreign counterpart. At the one-year horizon, we instead use the LIBOR interest-rate swap spread over the US Treasury note yield.[46]

III.B. Covered Interest Parity and the US Dollar Liquidity Premium

The primary variable we will use to capture the dollar premium, $\lambda_t^\$$, will be the LIBOR cross-currency basis—the deviation from covered interest parity among advanced country interbank borrowing rates—as we now explain.

Unlike UIP, covered interest parity (CIP) refers to a comparison of returns on debt instruments where exchange rate uncertainty is eliminated through the sale of one instrument's gross proceeds in the forward exchange market. An investment in a foreign currency debt instrument can effectively be transformed into a synthetic dollar investment if coupled with a forward exchange market sale of the foreign currency payoff, in which a counterparty agrees to exchange dollars for the foreign currency on the payoff date at a pre-agreed price (the forward exchange rate). CIP holds when synthetic dollar loans carry the same return or cost as comparable direct dollar loans. If f_t denotes the forward foreign currency price of dollars on date t, then in terms of our earlier notation, CIP holds when $i_t^L = i_t^{L*} + s_t - f_t$, or when:

$$i_t^{L*} = i_t^L + f_t - s_t. \tag{11}$$

Comparing equation (11) to equation (3) shows that UIP and CIP are equivalent if and only if $f_t = \mathbb{E}_t s_{t+1}$, but long-standing evidence firmly rejects that equality.

Indeed, CIP itself has failed to hold among different classes of low-risk or riskless bonds due to factors that are closely linked to exchange rate fluctuations. For market interest rates such as LIBOR, CIP deviations were small up through 2007–2008, but big and fairly persistent deviations from CIP have emerged since. Relative to the US dollar as the home currency,

46. Many empirical studies analyze LIBOR CIP, even though LIBORs are indicative and may not be perceived as absolutely risk-free in all circumstances. However, analysis based on even less risky rates such as the overnight indexed swap (OIS) rate yields similar conclusions (Du and Schreger 2022).

the gap $x_t^L \equiv i_t^{L*} - (i_t^L + f_t - s_t)$—called the LIBOR dollar basis—has generally been positive for most Group of Ten (G10) currencies since the global financial crisis, implying that $i_t^L < i_t^{L*} + s_t - f_t$: the cost of borrowing dollars directly is below that of synthetic dollar borrowing (for example, borrowing euros and selling them spot for dollars while simultaneously entering a forward contract to sell the dollars for euros upon maturity of the original euro loan).[47] In contrast, the Treasury basis, defined with respect to government bond rates (and with i_t denoting the US Treasury rate and i_t^* the foreign government bond rate) is $x_t \equiv i_t^* - (i_t + f_t - s_t)$. The condition $x_t = 0$ did not hold closely even before the financial crisis. It has not held afterward either, but x_t has become more closely correlated with x_t^L, which had a much smaller variance than x_t before the crisis but has had a generally similar variance since. Figure 13 illustrates the behavior of the two bases, for both the three-month and one-year investment horizons.

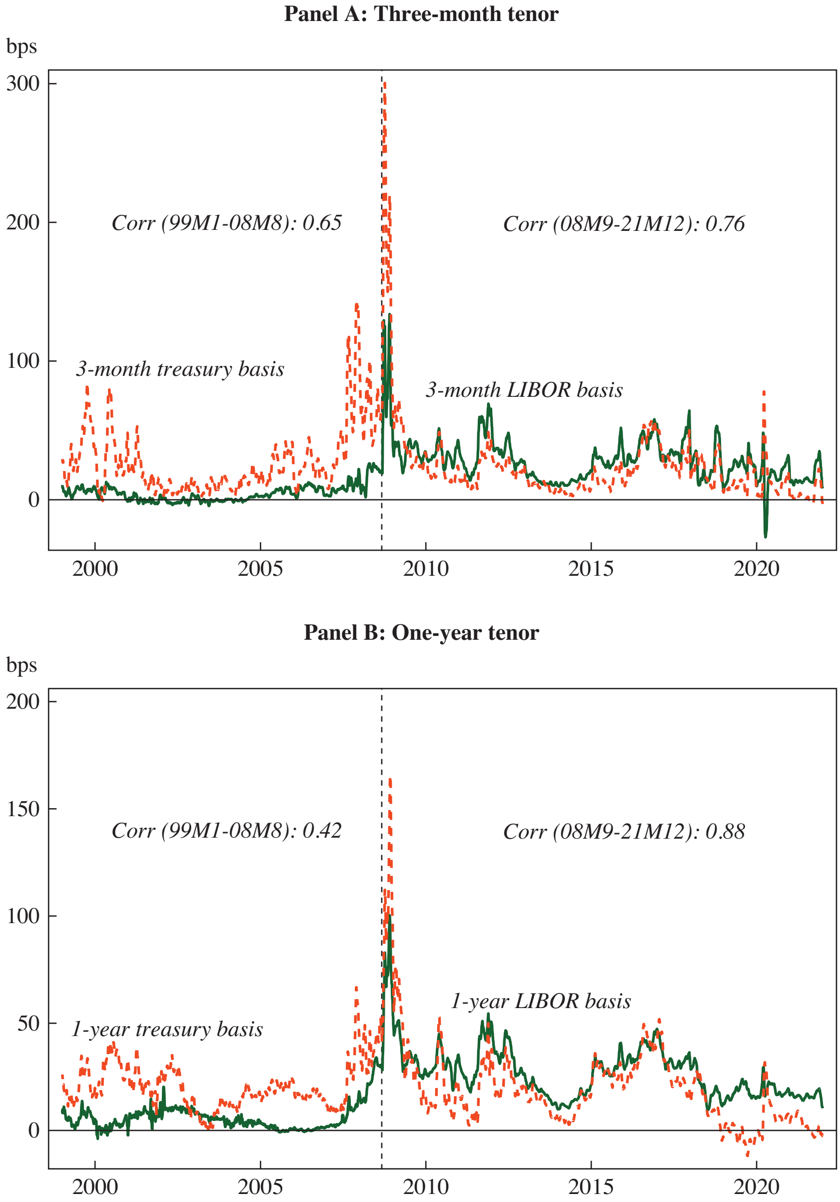
Du, Im, and Schreger (2018) have highlighted the Treasury premium as a measure of the relative convenience yield from holding US Treasury securities. Krishnamurthy and Lustig (2019), Jiang, Krishnamurthy, and Lustig (2021), and Engel and Wu (forthcoming) posit that Treasury basis fluctuations have a causal impact on dollar exchange rates. In those analyses, the advantage of US Treasury obligations arises from two (likely related) sources: the greater liquidity of Treasuries relative to privately issued bonds and the greater liquidity of dollar bonds relative to non-dollar bonds. But it is not straightforward to identify separately the two components of the convenience yield.

We have taken the relative spread $\gamma_t \equiv i_t^L - i_t - (i_t^{L*} - i_t^*)$ between private and central government issuers as a measure of the relative liquidity of US Treasuries. This measure, however, should bear little connection to the dollar's special international role, as the spreads it compares are for bonds of like currency denomination. Notice, however, that:

$$\begin{aligned}\gamma_t &= i_t^L - i_t - \left[i_t^{L*} + s_t - f_t - \left(i_t^* + s_t - f_t\right)\right]\\ &= i_t^L - \left(i_t^{L*} + s_t - f_t\right) - \left[i_t - \left(i_t^* + s_t - f_t\right)\right]\\ &= x_t - x_t^L,\end{aligned}$$

47. The US dollar basis has generally been negative for the Australian and New Zealand dollars, for reasons elucidated by Borio and others (2016) and Liao and Zhang (2020). For a broad discussion of the literature on deviations from CIP, see Du and Schreger (2022). Note that the literature generally defines the US dollar basis with a sign opposite to our convention. Given the wider scope of our discussion in this paper, however, we judged the definition in the text to be less confusing for readers.

Figure 13. LIBOR and Treasury Basis, 1999–2021



Sources: Bloomberg; Refinitiv.

Note: Ten-day moving average of daily deviations from CIP for three-month LIBOR rates and Treasury yields. Cross-sectional average is taken over CAD, CHF, DKK, EUR, GBP, JPY, NOK, and SEK. Vertical line marks September 2008. Pairwise correlations between the level of the average Treasury basis and the average LIBOR basis are computed and reported. One-year LIBOR bases are calculated based on LIBOR interest rate swaps.

which implies that:

$$x_t = x_t^L + \gamma_t. \tag{12}$$

Equation (1) is the key to our rationale for proxying $\lambda_t^{\$}$ by the LIBOR basis. As a first step, consider the thought experiment of a world with no financial frictions, in which markets would conduct full and efficient arbitrage between currencies in interbank markets. Because the assets involved in that arbitrage have identical liquidity characteristics apart from their currencies of denomination, any observed nonnegative dollar basis would have to reflect $\lambda_t^{\$}$. In that idealized world, equation (12) cleanly allocates the total Treasury premium between a component related dollar denomination per se and a component entirely due to the inherent comparative liquidity of Treasury obligations versus market-issued obligations. The main drivers of both $\lambda_t^{\$}$ and γ_t would be factors like global safe asset demand, risk aversion, and bond supplies that alter marginal convenience yields even with unconstrained intermediaries.[48]

Real-world financial markets are beset by trading constraints, however, and the LIBOR dollar basis therefore reflects not only the dollar's marginal liquidity value but also market frictions.[49] A range of evidence supports the link between intermediaries' balance sheet capacity and deviations from CIP, as discussed by Du (2019) and Du and Schreger (2022). Conversely, Federal Reserve swaps of dollars with foreign central banks, which lend the dollars to domestic banks with constrained alternative dollar access, have limited basis spreads by effectively filling in for scarce private balance sheet space (Bahaj and Reis 2022; Goldberg and Ravazzolo 2022). Notwithstanding the strong influence of market frictions on the dollar LIBOR basis, it still can serve as a stand-in for dollar liquidity in a regression equation for the dollar exchange rate that also controls for direct indicators of financial stress as well as the Treasury relative liquidity factor, γ_t.

48. In the interest arbitrage comparison, the combination of a cash position in a foreign asset and a forward purchase of dollars might inherit some fraction of the dollar convenience yield $\lambda_t^{\$}$, but as Jiang, Krishnamurthy, and Lustig (2021) argue, that fraction would most likely be strictly less than 1.

49. As we observed earlier, the convenience yields themselves are likely to depend partly on market frictions. Especially in the presence of frictions, the separability of US Treasury attributes one might be tempted to infer from the idealized version of equation (12) is implausible. For example, the depth of the US Treasury market surely enhances the value of "dollarness" for many other dollar-denominated assets.

Below, we will also consider the Treasury basis x_t as a single regressor in place of the LIBOR basis and γ_t, as Krishnamurthy and Lustig (2019), Jiang, Krishnamurthy, and Lustig (2021), and Engel and Wu (forthcoming) do. According to equation (12), the Treasury basis is the sum of the LIBOR dollar basis and γ_t, so in principle it could serve as an indicator of both those convenience yields if they are weighted equally by investors. However, there is no reason to assume that equal weighting holds, and our baseline specification with both x_t^L and γ_t does not do so. The data support that approach.[50]

It is well known that the LIBOR basis (like the Treasury basis) is closely associated with the dollar: dollar appreciations correspond to a wider basis.[51] This correlation admits different channels of causation. It may be that the basis-dollar link mainly reflects shifts in global investor preferences or asset supplies that drive the dollar, perhaps through a convenience yield channel. But a complementary account holds that dollar movements reflect shifts in global financial conditions that simultaneously alter financial intermediaries' balance sheet space and thereby their propensities to arbitrage return gaps via the forward exchange market.[52] The relationship between global balance sheet capacity and the dollar owes to more than just common risk aversion or safe asset demand shocks. Through an additional feedback loop, dollar appreciation, whatever its cause, itself impairs the balance sheets of unhedged dollar debtors, tightening financial conditions and widening US dollar bases. These possibilities all dictate caution in interpreting the exchange rate regressions that we present next. At best, they capture key correlations that are potentially indicative of alternative causal mechanisms.

50. In unreported estimates, we find that when we enter both the Treasury basis x_t and γ_t in the regression, the estimated coefficient of γ_t is negative and smaller in absolute value than the estimated coefficient of x_t, which itself is the same as the estimated coefficient of x_t^L in our baseline regressions. On the other hand, as our findings below show, the estimated coefficient of x_t, when entered alone without γ_t, is biased downward owing to omitted variable bias from leaving out γ_t. These patterns are consistent with the assumption that x_t^L and γ_t indeed capture different components of the Treasury liquidity yield, but with the pure dollar effect $\lambda_t^{\$}$ quantitatively more important to investors on average over the entire sample period.

51. See, for example, Avdjiev, Du and others (2019) and Cerutti, Obstfeld, and Zhou (2021).

52. Du (2019) makes this argument, also documenting the closer co-movement between the LIBOR and Treasury bases after the global financial crisis (see figure 13). That co-movement suggests a relatively larger role for $\lambda_t^{\$}$ after the crisis and for γ_t before. The substantial correlation coefficient of the two bases before the crisis, however, suggests a significant role for $\lambda_t^{\$}$ even then.

III.C. Empirical Exchange Rate Equations

We next present and discuss the results of estimating equations (8)–(10) by ordinary least squares, using a monthly panel of G10 currencies starting in 1999. As discussed in the previous sections, for each specification, we present estimates for three-month and one-year changes in the log nominal end-of-period bilateral exchange rate of G10 currencies against the dollar, including currency fixed effects throughout. As overlapping samples are used, we report heteroskedasticity-robust and autocorrelation-robust standard errors (Driscoll and Kraay 1998). Three-month log changes are measured at an annual rate. Further details on the data are in online appendix B.

In each of tables 2–4, the first two columns estimate over 1999–2021 and the second two estimate over the post-crisis period 2010–2021. Odd-numbered columns report equations with the LIBOR basis x_t^L and γ_t both included, while even-numbered columns instead include the Treasury basis x_t as the sole convenience yield proxy. In the estimation, all interest rates regardless of tenor are expressed as annualized rates.

Panels A and B of table 2 report estimates of equation (8). The two panels are based, respectively, on three-month and one-year exchange rate changes, and three-month and one-year changes in three-month and one-year interest rates. Over all specifications and samples, the change in the three-month US Treasury interest rate relative to the foreign bond rate is highly economically and statistically significant. For example, column 1 in panel A implies that a 10 basis point increase in the annualized three-month Treasury differential over a quarter appreciates the dollar by 125.08/4 = 31.3 basis points over that quarter. The same column in panel B implies that a 10 basis point rise in the one-year Treasury differential over a year appreciates the dollar by 40.7 basis points.

In all regressions the lagged real exchange rate is also highly significant, with real appreciation predicting nominal depreciation over the following period. This mean reversion, though estimated fairly precisely over the entire sample, is rather gradual (generally around 2–4 basis points depreciation of the foreign currency per year for a 10 basis point real appreciation of the dollar), in line with the copious evidence of slow mean reversion in real exchange rates (Itskhoki 2021). Estimated mean reversion is higher over the post-crisis sample.

Turning to indicators associated with the convenience yield of dollar Treasuries, in odd-numbered columns of both panels of table 2, the γ_t variable measuring the relative liquidity of Treasuries (apart from their

Table 2. Exchange Rate Equations: Short-Term Rates

	Δ = *three months; fc quarter-over-quarter depreciation*			
Variables	*(1) 1999–2021*	*(2) 1999–2021*	*(3) 2010–2021*	*(4) 2010–2021*
Panel A: Three-month horizon				
$\Delta(i^{US}_{3m,t} - i^{*}_{3m,t})$	12.508***	13.313***	15.749***	17.340***
	(2.772)	(2.751)	(4.641)	(3.989)
$\Delta\gamma_{3m,t}$	2.990		4.706	
	(3.214)		(5.756)	
Δ three-month LIBOR basis (pp)	10.093***		11.080**	
	(2.776)		(4.797)	
Δ three-month Treasury basis (pp)		6.274***		8.877**
		(2.402)		(3.445)
Δ log VIX	0.052	0.052	0.085**	0.086**
	(0.034)	(0.034)	(0.041)	(0.041)
Δ excess bond premium	17.701***	17.400***	15.058***	14.263***
	(3.454)	(3.382)	(4.610)	(4.457)
Lag RER	–0.198***	–0.211***	–0.448***	–0.447***
	(0.070)	(0.072)	(0.078)	(0.079)
Observations	2,757	2,757	1,440	1,440
Adjusted R^2	0.252	0.250	0.220	0.219
Currency FE	✓	✓	✓	✓
Lagged controls	✓	✓	✓	✓
Driscoll and Kraay (1998) lags	3	3	3	3

currency denomination) is correctly signed but statistically insignificant. The LIBOR basis has the theoretically correct sign and is quite significant for three-month changes. The estimated coefficient of the Treasury basis is smaller than that of the LIBOR basis over both estimation samples, owing to the former's conflation of the dollar effect $\lambda^{\$}_t$ with the weaker effect γ_t. In panel B for one-year exchange rate changes, both dollar bases have correct signs but generally lower statistical significance than in panel A. Only for the post-crisis sample do we find statistically significant coefficients (at the 5 percent level) associated with both bases. The coefficients of the LIBOR basis are comparable to those of interest rates, if usually somewhat smaller.

Next consider the two regressors meant to capture financial market stresses. At the three-month horizon (panel A), the influence of the VIX has the expected sign but is very small, with a 10 basis point increase in the index corresponding to a minuscule 0.5/4 = 0.125 basis point appreciation of the dollar over the quarter for the entire sample and just below 0.9/4 = 0.225 basis point post-crisis. Neither estimate is significant at the 5 percent level. However, the EBP variable is highly statistically significant

Table 2. Exchange Rate Equations: Short-Term Rates *(Continued)*

	Δ = one year; fc year-over-year depreciation			
Variables	*(1) 1999–2021*	*(2) 1999–2021*	*(3) 2010–2021*	*(4) 2010–2021*
Panel B: One-year horizon				
$\Delta(i^{US}_{1y,t} - i^{*}_{1y,t})$	4.069***	4.043***	4.069***	4.062***
	(1.060)	(1.063)	(1.103)	(1.168)
$\Delta\gamma_{1y,t}$	2.252		4.080	
	(1.917)		(3.067)	
Δ one-year LIBOR basis (pp)	2.807		8.102**	
	(2.604)		(3.392)	
Δ one-year Treasury basis (pp)		2.621		5.595**
		(1.794)		(2.738)
Δ log VIX	−0.024	−0.023	−0.003	−0.003
	(0.019)	(0.020)	(0.023)	(0.024)
Δ excess bond premium	7.534***	7.490***	6.143***	6.301***
	(1.205)	(1.223)	(1.722)	(1.861)
Lag RER	−0.205***	−0.200***	−0.386***	−0.383***
	(0.044)	(0.044)	(0.049)	(0.052)
Observations	2,725	2,742	1,440	1,440
Adjusted R^2	0.449	0.447	0.489	0.476
Currency FE	✓	✓	✓	✓
Lagged controls	✓	✓	✓	✓
Driscoll and Kraay (1998) lags	12	12	12	12

Source: Authors' calculations.

Note: Table reports the results of estimating equation (8) on a monthly sample for bilateral exchange rates of G10 currencies against the US dollar. Spot exchange rates are expressed in units of foreign currency per US dollar. The variables $\Delta\gamma_{3m,t}$ and $\Delta\gamma_{1y,t}$ are the relative spread difference between US and foreign three-month LIBOR rates and one-year LIBOR swap rates, respectively, against yields on government securities of like tenor. The Treasury basis at tenor j is defined as $i^{*}_{j,t} - (i^{US}_{j,t} + f_{j,t} - s_t)$, where f and s are forward and spot exchange rates. For panel A, overlapping quarterly changes along with interest rates and bases at three-month tenors are used. The dependent variable is the *annualized* quarter-over-quarter depreciation rate. For panel B, overlapping yearly changes and depreciation rates are used. All variables are expressed in percentages (or in 100 times log terms). The table reports standard errors per Driscoll and Kraay (1998).

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

with a large coefficient. In column 1 of panel A, a 10 basis point rise in the EBP is associated with a currency appreciation over the quarter of 177/4 ≈ 44 basis points, with a slightly smaller correlation in column 3. The estimated coefficient of EBP is only slightly lower post-crisis, and it remains statistically significant at the 1 percent level.[53]

53. In standard deviation terms, a one standard deviation increase in 100 × log VIX translates into a 4.4 basis point dollar appreciation over the same quarter, based on estimation over the entire 1999–2021 sample. A one standard deviation increase in EBP is associated with a 31 basis point dollar appreciation over the same horizon and sample. The corresponding numbers post-crisis are 7.2 basis points (for the VIX) and 11.8 basis points (for EBP).

Table 3. Exchange Rate Equations: Long-Term Rates, Short-Term Liquidity Premium

	Δ = three months; fc quarter-over-quarter depreciation			
Variables	*(1) 1999–2021*	*(2) 1999–2021*	*(3) 2010–2021*	*(4) 2010–2021*
Panel A: Three-month horizon				
$\Delta(i^{US}_{10y,t} - i^{*}_{10y,t})$	38.975***	39.928***	42.666***	44.166***
	(4.190)	(4.112)	(5.579)	(5.233)
$\Delta(tp^{US}_{10y,t} - tp^{*}_{10y,t})$	−23.773***	−24.609***	−25.408***	−26.635***
	(3.882)	(3.762)	(5.334)	(5.134)
$\Delta\gamma_{3m,t}$	−1.933		−0.189	
	(2.804)		(4.687)	
Δ three-month LIBOR basis (pp)	5.325**		9.792**	
	(2.382)		(4.313)	
Δ three-month Treasury basis (pp)		1.307		5.914*
		(1.875)		(3.046)
Δ log VIX	0.074**	0.074**	0.113***	0.112***
	(0.036)	(0.036)	(0.043)	(0.042)
Δ excess bond premium	20.091***	19.729***	16.956***	15.770***
	(2.795)	(2.683)	(4.072)	(3.755)
Lag RER	−0.177***	−0.188***	−0.422***	−0.421***
	(0.060)	(0.061)	(0.072)	(0.072)
Observations	2,757	2,757	1,440	1,440
Adjusted R^2	0.350	0.348	0.338	0.334
Currency FE	✓	✓	✓	✓
Lagged controls	✓	✓	✓	✓
Driscoll and Kraay (1998) lags	3	3	3	3

Panel B of table 2 indicates that the VIX has the wrong sign (but is insignificant) for one-year exchange rate changes. The excess bond premium is sizable and significant in panel B in all specifications, with an even stronger influence than in panel A. In every column of panel B, a 10 basis point rise in EBP is estimated to appreciate the currency by more than 60 basis points over the year—at least 1.5 times the association with a 10 basis point rise in the interest differential.

Finally, the R^2 coefficients are notable. In the equation estimates that panel A reports, all R^2s are between 0.2 and 0.3. In panel B, however, R^2s fall between 0.4 and 0.5. Taken together, the variables in the regressions have considerable explanatory power for contemporaneous year-to-year exchange rate changes.

Table 3 reports estimates of equation (9). As expected, estimated coefficients for changes in long-term interest differentials are much larger than for short-term differentials, which in equation (8) stand in for news about

Table 3. Exchange Rate Equations: Long-Term Rates, Short-Term Liquidity Premium *(Continued)*

	Δ = one year; fc year-over-year depreciation			
Variables	*(1)* *1999–2021*	*(2)* *1999–2021*	*(3)* *2010–2021*	*(4)* *2010–2021*
Panel B: One-year horizon				
$\Delta(i^{US}_{10y,t} - i^{*}_{10y,t})$	9.614***	9.625***	9.341***	9.470***
	(2.138)	(2.144)	(1.812)	(2.042)
$\Delta(tp^{US}_{10y,t} - tp^{*}_{10y,t})$	–6.034***	–6.014***	–6.744***	–5.782***
	(2.007)	(2.024)	(2.073)	(2.042)
$\Delta\gamma_{1y,t}$	0.886		3.390	
	(2.184)		(3.598)	
Δ three-month LIBOR basis (pp)	0.148		9.694***	
	(3.001)		(3.593)	
Δ three-month Treasury basis (pp)		0.778		5.636*
		(1.994)		(3.042)
Δ log VIX	–0.013	–0.012	0.004	0.008
	(0.019)	(0.020)	(0.021)	(0.020)
Δ excess bond premium	7.866***	7.844***	6.655***	6.848***
	(1.349)	(1.382)	(1.583)	(1.706)
Lag RER	–0.192***	–0.185***	–0.366***	–0.362***
	(0.043)	(0.044)	(0.052)	(0.055)
Observations	2,725	2,742	1,440	1,440
Adjusted R^2	0.462	0.461	0.554	0.533
Currency FE	✓	✓	✓	✓
Lagged controls	✓	✓	✓	✓
Driscoll and Kraay (1998) lags	12	12	12	12

Source: Authors' calculations.

Note: Table reports the results of estimating equation (9) on a monthly sample for bilateral exchange rates of G10 currencies against the US dollar. Spot exchange rates are expressed in units of foreign currency per US dollar. The term premium differential, $tp^{US}_{10y,t} - tp^{*}_{10y,t}$ is estimated based on zero-coupon government bond yield curves from Bloomberg and national central banks, using the model of Adrian, Crump, and Moench (2013) with four principal components of yields as the state variables. The variables $\Delta\gamma_{3m,t}$ and $\Delta\gamma_{1y,t}$ are the relative spread difference between US and foreign three-month LIBOR rates and one-year LIBOR swap rates, respectively, against yields on government securities of like tenor. The Treasury basis at tenor j is defined as $i^{*}_{j,t} - (i^{US}_{j,t} + f_{j,t} - s_t)$, where f and s are forward and spot exchange rates. For panel A, overlapping quarterly changes along with interest rates and bases at three-month tenors are used. The dependent variable is the annualized quarter-over-quarter depreciation rate. For panel B, overlapping yearly changes and depreciation are used. All variables are expressed in percentages (or in 100 times log terms). The table reports standard errors per Driscoll and Kraay (1998).

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

future short-term interest rates. In panel A, column 1, a 10 basis point rise in the ten-year yield differential in favor of Treasuries is associated with a 389.75/4 ≈ 97 basis point appreciation of the dollar over the same quarter. The association is somewhat stronger in the quantitative easing (QE) era following the financial crisis. In panel B, column 1, a 10 basis point rise in the ten-year Treasury yield differential is associated with a 96 basis point dollar appreciation over the same year. The coefficient is roughly stable across specifications and periods in panel B. In all table 3 estimates, the term premium differential has the negative sign that equation (9) implies, but the absolute sizes of its coefficients are smaller than those for long-term interest differentials, contrary to the theory. This pattern may reflect that the term premium variables are estimated, and therefore measured with error. Throughout table 3, the estimated role of the lagged real exchange rate conforms to the pattern in table 2.

The change in γ_t is statistically insignificant in all cases, but the coefficients for the LIBOR basis are of correct sign and statistically significant at the 5 percent level or better, except in column 1 of panel B. In column 3 of panel B, covering post-crisis data, the variable's estimated coefficient is similar to that of the long-term interest differential. On the other hand, EBP is statistically significant and sizable for all specifications and time periods. The VIX index is now statistically significant in panel A for three-month changes, but its coefficient remains small in magnitude and is not ever statistically significant for the longer horizon (one-year, panel B). The R^2 coefficients are higher across the board than in table 2, reaching the range of 0.46–0.56 in panel B.

The strong estimated relationship of long-term interest differentials with exchange rates and the impressive in-sample fit of exchange rate equations based on long-term rates is consistent with recent theories of debt-driven exchange rate movements such as Greenwood and others (2020) and Gourinchas, Ray, and Vayanos (2022), as well as with several econometric studies on the effects of QE by major central banks, such as Dedola and others (2021). In equation (9), however, long-term rate differentials are entered jointly with the term premium, their difference standing in for the expected sum of future short-term rate differentials. Furthermore, the term premium is measured with error. A better sense of the impact of long-term rates may come from estimates of equation (10), in which the role of long-term rates follows directly from potential arbitrage among long-term government yields.

Table 4 presents estimates of that equation. The regressions in this table construct γ_t and cross-currency bases using ten-year LIBOR interest rate

Table 4. Exchange Rate Equations: Long-Term Rates, Long-Term Liquidity Premium

	Δ = three months; fc quarter-over-quarter depreciation			
Variables	*(1) 1999–2021*	*(2) 1999–2021*	*(3) 2010–2021*	*(4) 2010–2021*
Panel A: Three-month horizon				
$\Delta(i^{US}_{10y,t} - i^{*}_{10y,t})$	26.918***	24.342***	24.332***	24.108***
	(3.337)	(3.310)	(3.296)	(3.375)
$\Delta\gamma_{10y,t}$	32.875***		19.870**	
	(4.382)		(9.511)	
Δ ten-year LIBOR basis (pp)	53.786***		49.961***	
	(10.316)		(14.653)	
Δ ten-year Treasury basis (pp)		33.648***		23.271**
		(5.188)		(9.085)
Δ log VIX	0.063*	0.071*	0.097**	0.106**
	(0.037)	(0.038)	(0.045)	(0.048)
Δ excess bond premium	16.411***	16.598***	14.430***	15.390***
	(2.947)	(2.869)	(3.994)	(4.063)
Lag RER	–0.154**	–0.154**	–0.334***	–0.335***
	(0.072)	(0.066)	(0.066)	(0.069)
Observations	2,695	2,727	1,440	1,440
Adjusted R^2	0.312	0.294	0.289	0.273
Currency FE	✓	✓	✓	✓
Lagged controls	✓	✓	✓	✓
Driscoll and Kraay (1998) lags	3	3	3	3

swaps (based on three-month float-to-float exchanges), as in Du, Tepper, and Verdelhan (2018). All four columns of panel A suggest that a 10 basis point rise in the ten-year Treasury yield differential correlates with a substantial dollar appreciation over the same quarter of about 250/4 = 62.5 basis points. For one-year changes (panel B), the association is higher over the entire sample (around a 94 basis point appreciation for a 10 basis point yield difference) but closer to panel A over the post-crisis sample (roughly a 75 basis point effect).

Liquidity differences between long-term government bond γ_t are influential on exchange rate movements. All estimates are significant at least at the 10 percent level in table 4. The statistical significance is weakest during the post-crisis subperiod for one-year exchange rate changes. The LIBOR basis is again statistically and economically extremely significant, with estimated coefficients well in excess of long-term interest gaps. Treasury bases have similar significance, as in all the tables, but with downward-biased coefficients. The VIX roughly follows the pattern of table 3, relevant for three-month exchange rate changes but small in magnitude and unimportant

Table 4. Exchange Rate Equations: Long-Term Rates, Long-Term Liquidity Premium *(Continued)*

	Δ = one year; fc year-over-year depreciation			
Variables	*(1) 1999–2021*	*(2) 1999–2021*	*(3) 2010–2021*	*(4) 2010–2021*
Panel B: One-year horizon				
$\Delta(i^{US}_{10y,t} - i^{*}_{10y,t})$	9.421***	8.153***	7.486***	7.589***
	(1.601)	(1.616)	(1.316)	(1.438)
$\Delta\gamma_{10y,t}$	10.159***		4.846*	
	(2.571)		(2.813)	
Δ ten-year LIBOR basis (pp)	16.031***		17.810***	
	(4.879)		(5.018)	
Δ ten-year Treasury basis (pp)		10.952***		8.227**
		(2.566)		(3.305)
Δ log VIX	0.002	−0.006	0.018	0.018
	(0.019)	(0.018)	(0.017)	(0.017)
Δ excess bond premium	6.721***	6.918***	5.846***	7.625***
	(1.173)	(1.243)	(1.416)	(1.655)
Lag RER	−0.197***	−0.203***	−0.371***	−0.371***
	(0.048)	(0.042)	(0.056)	(0.060)
Observations	2,624	2,673	1,440	1,440
Adjusted R^2	0.472	0.446	0.510	0.485
Currency FE	✓	✓	✓	✓
Lagged controls	✓	✓	✓	✓
Driscoll and Kraay (1998) lags	12	12	12	12

Source: Authors' calculations.

Note: Table reports the results of estimating equation (10) on a monthly sample for bilateral exchange rates of G10 currencies against the US dollar. Spot exchange rates are expressed in units of foreign currency per US dollar. The variable $\Delta\gamma_{10y,t}$ is the relative spread difference between US and foreign ten-year LIBOR swap rates against yields on government securities of like tenor. For panel A, overlapping quarterly changes along with interest rates and bases at three-month tenors are used. The dependent variable is the annualized quarter-over-quarter depreciation rate. For panel B, overlapping yearly changes and depreciation are used. All variables are expressed in percentages (or in 100 times log terms). The table reports standard errors per Driscoll and Kraay (1998).

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

for one-year changes. Also consistent with the other tables, EBP remains highly significant and strongly associated with both one-quarter and one-year exchange rate movements. The R^2s are slightly lower than in table 3, albeit still sizable.

To summarize the results of tables 2–4, US Treasury interest rate differentials are important correlates of dollar exchange rate changes, but long-term yield differentials are especially powerful over our entire sample period and since the global financial crisis. These correlations indicate the importance of monetary and debt management policies. Other factors, however, play important roles, in line with the recent literature on exchange rate determination. One such factor is the cross-currency dollar basis—LIBOR or Treasury—with the former being a more direct measure of the specific liquidity value of the US dollar to global investors. While both bases reflect the marginal liquidity advantage of US Treasury obligations as seen by market participants, and therefore also monetary and debt policies, they also reflect global safe asset demand and related financial market frictions. In risk-off market episodes, the demand for safe dollar assets rises while financial intermediary constraints simultaneously tighten. One widely monitored index of risk sentiment, the VIX, has some contemporaneous correlation with the dollar exchange rate in the short term (over three months) but nothing detectable at longer term (over a year). While we find the LIBOR basis to have a strong and highly statistically significant correlation with the dollar, the most consistently influential correlate (aside from interest rates themselves) is the EBP (Gilchrist and Zakrajšek 2012)—an indicator of credit market sentiment. This finding provides strong evidence that US financial conditions, alongside monetary policies, are key factors influencing the dollar and potentially the global financial cycle.

III.D. EBP Shocks and Emerging Markets

The exchange rate equations we estimated in the previous section illustrate the important connection between dollar movements and US financial conditions. The high and consistent correlation of EBP movements with dollar shocks invites a direct look at how EBP shocks themselves affect emerging market economies. The EBP is based on US data and is a strong predictor of US recessions, but it could also capture broader global movements in risk appetite and financial conditions. In this section we return to the LP framework of section II and show that EBP shocks predict sharp contractions in emerging market economies. Section II reported the average results of "generic" dollar shocks, possibly driven by a range of factors including the EBP, but here we home in on the specific role of EBP shocks, as have a number of other recent studies.[54] To that end, we replace the

54. Ben Zeev (2019) conducts an exercise similar to ours but focusing on the state-dependent response of EMDEs to EBP shocks according to whether the exchange rate is fixed. Cesa-Bianchi and Sokol (2022) study the transmission of EBP shocks from the United States to the United Kingdom. Gilchrist and others (2022) study how several proxies for global risk affect sovereign spreads on dollar-denominated bonds. They find that the EBP has the strongest influence on spreads. Georgiadis, Müller, and Schumann's (2021) counterfactual analysis based on a Bayesian vector autoregression framework suggests that dollar appreciation significantly amplifies the contractionary effect of global risk shocks, mostly through a financial channel.

contemporaneous dollar appreciation shock in equation (1) with quarterly EBP changes, while keeping the lagged change in the nominal AE dollar index in the same forecasting equation to control for any lagged dollar impact on EMDE variables not captured by the EBP.

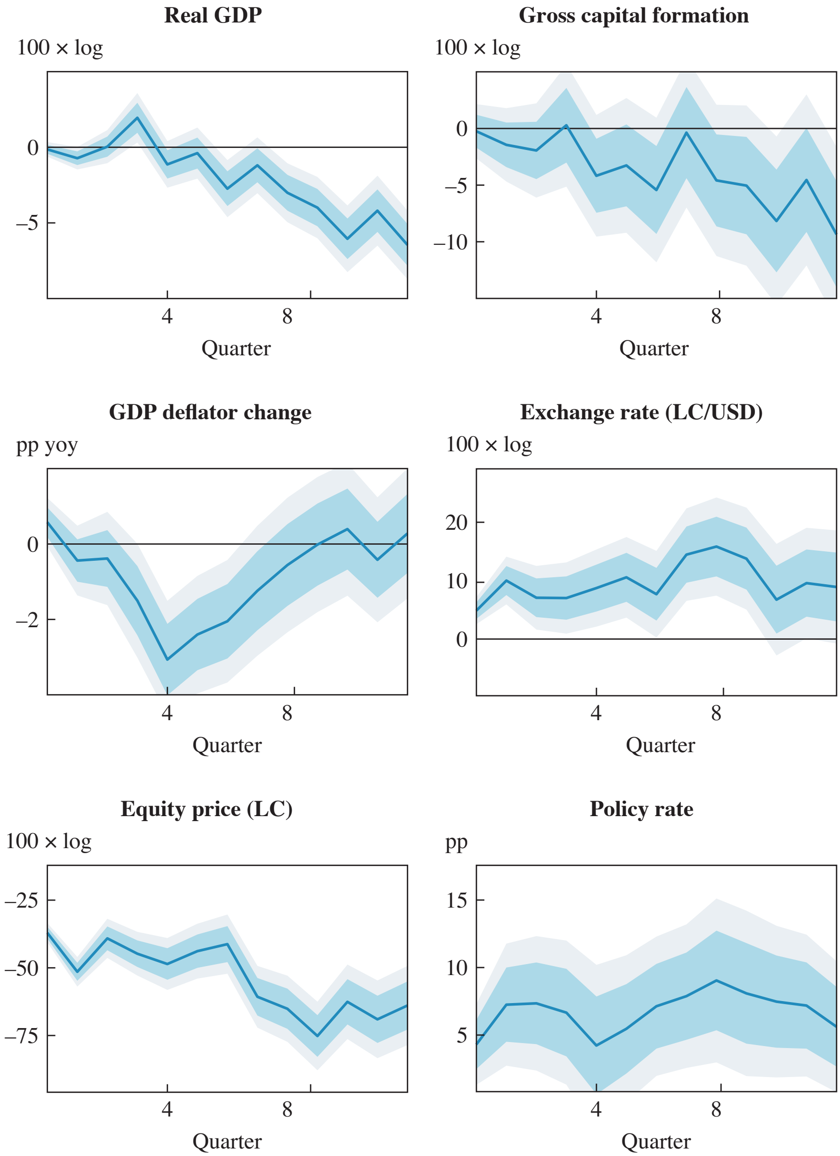
Figure 14 plots selected impulse responses of EMDE economic and financial variables to a 250 basis point increase in EBP. We find an overwhelmingly contractionary impact as in section II, but the slump seems to gather strength more slowly and then becomes deeper and more persistent than the one caused by a general dollar shock. Real output contracts below trend by a cumulative 5 percentage points after ten quarters, driven by steep declines in consumption and investment that more than offset a rise in net exports (see online appendix A.3 for impulse responses not included in the figure). The peak exchange rate depreciation against the dollar exceeds 10 percent, accompanied by worsening terms of trade and an overall contraction in trade volumes. The shock also has a deflationary impact on both domestic and trade-related prices. Looking at financial variables, nominal credit shrinks. The policy rate jumps upward by nearly 5 percentage points on impact (and peaks at 10 percentage points) while dollar borrowing costs, proxied by the EMBI spread, rise on impact and the domestic equity prices enter a prolonged decline. While US dollar appreciation is generally negative for EMDEs' economic health, dollar movements associated with the risk appetite shifts that EBP captures have especially severe impacts.

IV. The Dollar's Unsettled Future

The start of the COVID-19 pandemic in the first quarter of 2020 saw panic in global financial markets, large financial capital outflows from EMDEs, and a sharp rise in the dollar. The US Treasury market itself became illiquid as a "dash for cash" developed in March. The global dollar cycle went sharply into contraction.

Central banks around the world made deep cuts to interest rates, and governments deployed aggressive fiscal support of their economies. Given the central role of US financial markets and the dollar, Federal Reserve actions were especially important in stabilizing world financial markets. Expansion of Federal Reserve swap lines and establishment of the Foreign and International Monetary Authorities (FIMA) Repo Facility—which ensured a buyer of last resort for foreign central banks desiring to sell US Treasury reserves—were central to the turnaround (Goldberg and Ravazzolo 2022). So were the Federal Reserve's renewed large-scale asset purchases and lending to the private sector, unprecedented in volume and scope.

Figure 14. Impulse Response: 2.5 Percent Increase of Excess Bond Premium



Source: Authors' calculations.

Note: The impulse response functions of EMDE economic and financial variables to a 2.5 percent increase in the EBP (Gilchrist and Zakrajšek 2012). Estimates are derived from the local projection, equation (1), but with the change in the dollar index against AE currencies replaced by quarterly changes in EBP. For regressions involving the GDP deflator, country-quarter observations with a year-over-year change greater than 50 percent are dropped. Equity prices are local currency stock market indexes. Heteroskedasticity-robust 90 percent and 68 percent confidence bands are reported.

Capital flowed back into EMDEs, the dollar retreated, and a new expansive stage of the global dollar cycle began (see figure 15).

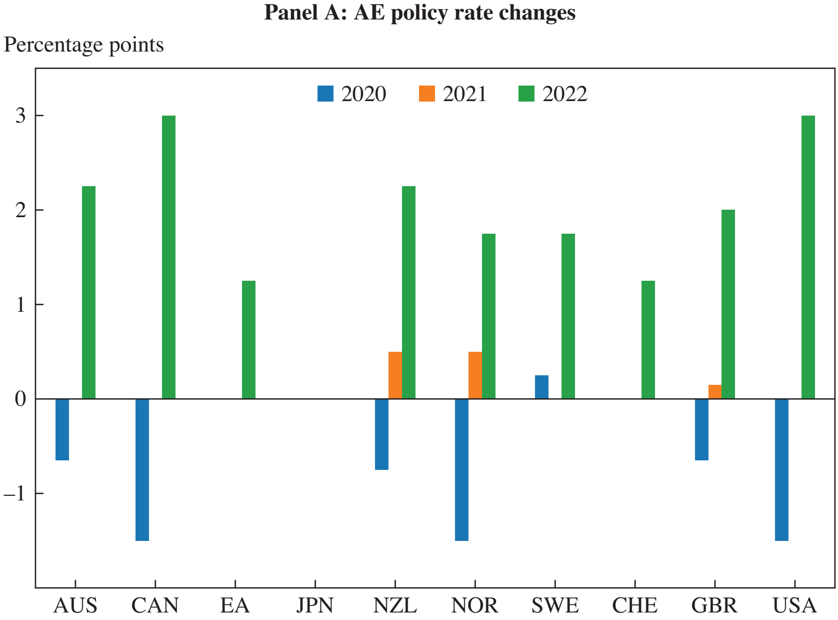
As the world economy reopened from pandemic lockdowns, demand pressures collided with supply constraints to generate a worldwide upsurge in inflation. The contribution of aggregate demand to inflation has been particularly high in the United States. Yet, while many EMDE central banks and a small number of AE central banks began raising policy interest rates in 2021 (panels A and B of figure 15), the Federal Reserve has been late to the game, raising the federal funds target by 25 basis points in March 2022 before scrambling to add another 50 basis points in May, 75 in June, 75 in July, and 75 more in September as US core inflation continued to rise.[55] As of this writing, two more 75 basis point hikes seem very possible in 2022. The result has been a sharp dollar appreciation, starting in mid-2021 when it became evident that faster US inflation would force the Federal Reserve to tighten earlier than markets had expected (panels C and D of figure 15). Now, a renewed contractionary phase of the global dollar cycle is underway. The effects will be economically harmful for many EMDEs, where both public- and business-sector debt loads rose significantly due to the pandemic. EMDEs will suffer as depreciation of their currencies raises the real value of dollar debts, as higher interest rates raise debt servicing burdens, and as slower growth erodes government tax receipts and business profits.

Indeed, EMDEs are facing a twofold challenge under current macroeconomic conditions. After making impressive progress to contain inflation over recent decades, they are raising domestic interest rates to prevent inflation from again becoming entrenched in the face of domestic currency depreciation and higher global commodity prices. At the same time, tighter financial conditions are having a contractionary effect, impairing balance sheets and worsening debt burdens.

An important research priority is to study exactly how EMDEs use their policy tools to cope with external financial shocks and whether these responses successfully reduce negative domestic repercussions. The macro tools deployed comprise monetary policy, foreign exchange intervention, fiscal policy, macroprudential policy, and direct measures to limit capital inflows and outflows. In particular, what is the role of the exchange rate—does it enable a more countercyclical response and otherwise buffer foreign shocks, as the results of this paper and others suggest, or is it a net shock

55. Board of Governors of the Federal Reserve System, "Policy Tools: Open Market Operations," https://www.federalreserve.gov/monetarypolicy/openmarket.htm.

Figure 15. Monetary Policies, Global Inflation, and the US Dollar Exchange Rate, 2020–2022



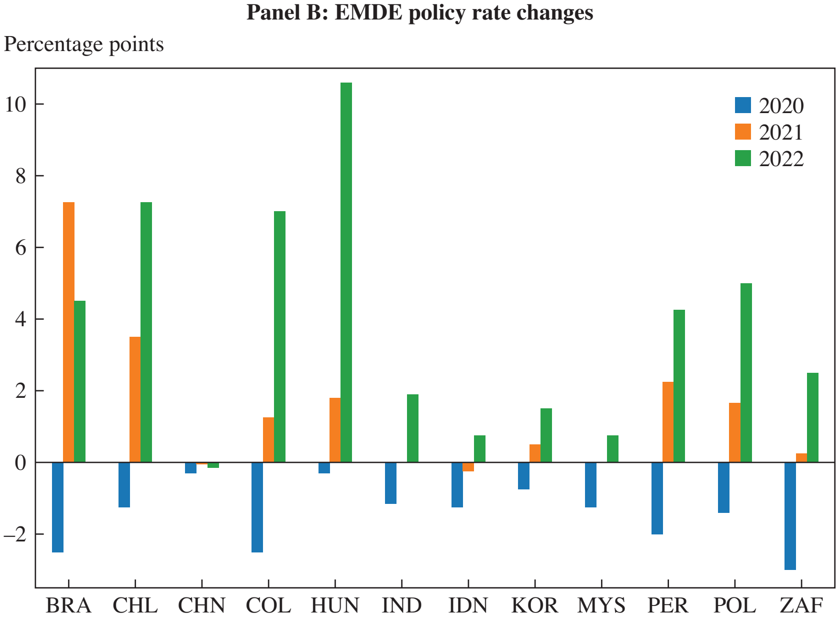
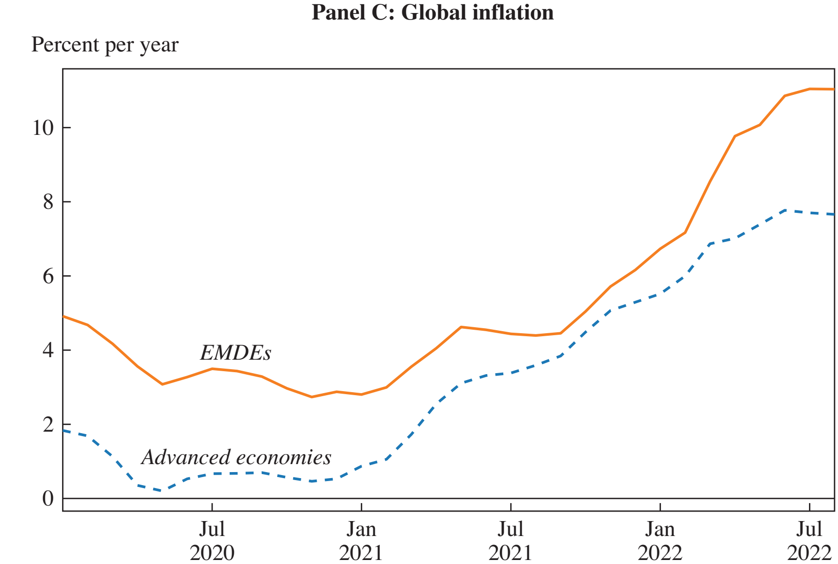
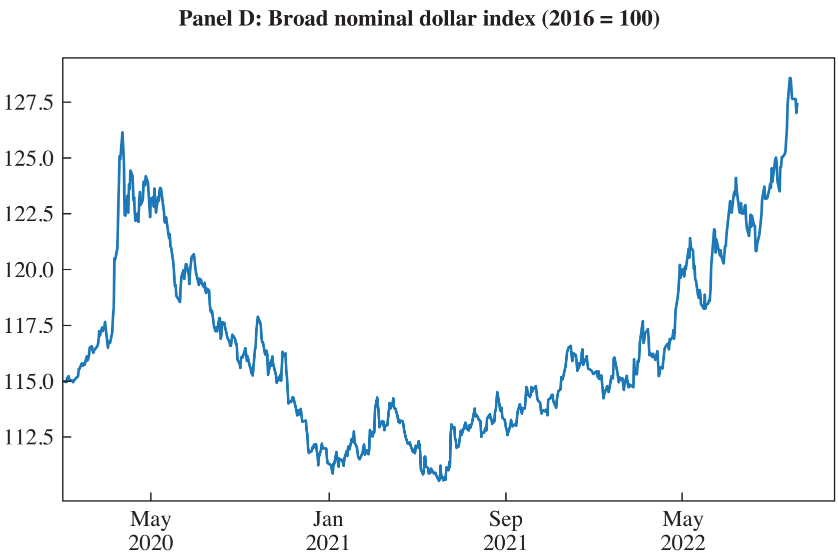


Figure 15. Monetary Policies, Global Inflation, and the US Dollar Exchange Rate, 2020–2022 *(Continued)*





Sources: Bank for International Settlements; IMF International Financial Statistics (via Haver); Refinitiv; Federal Reserve H.10 release.

Note: Panels A and B plot year-over-year and year-to-date for 2022 changes in policy interest rates for a set of AEs and emerging market economies. For 2022, the latest observations on policy rates were retrieved on October 20, 2022. EA in panel A refers to euro area (European Central Bank main refinancing rate). Panel C shows monthly values of year-over-year CPI inflation at an annual rate for both the G10 AEs and fifty-one EMDEs. Inflation rates are group weighted averages with 2015 nominal GDP weights.

amplifier? What are the transmission channels of currency changes and how important are they quantitatively in different countries? In a recent survey of emerging market central banks by the Committee on the Global Financial System (2021, 71), only seven of eighteen agreed that local currency depreciation is expansionary, while two believed it was contractionary and nine simply did not respond to the question. Perhaps the nonresponses reflected the question's failure to specify the shock driving local depreciation—a critical consideration. The results of this paper support the proposition that regimes with some exchange rate flexibility, central bank credibility, and lower foreign currency liabilities are helpful as platforms for effective EMDE policy responses to shocks. The current dollar cycle will retest the resilience of EMDE policy frameworks that in general were effective in coping with the COVID-19 shock early in 2020. This time, the test occurs in an environment of elevated inflation and rising, not falling, global interest rates.

What policy options do EMDEs have in their current situation? Those that are available may have limited effectiveness and come with significant trade-offs, though some EMDEs are already pursuing them. One option is foreign exchange intervention, that is, sales of hard currency reserves (mostly dollars) for the domestic currency, aimed at resisting its depreciation. This approach could in principle allow somewhat stronger currencies and lower policy interest rates consistent with less imported inflation. However, many EMDEs rely on sizable reserve war chests to inspire market confidence, and they could burn through large volumes of their holdings in prolonged battles against a strong dollar. If advanced country central banks were to extend their swap line offerings, that would effectively bolster EMDE foreign exchange reserves.

A second approach would be to moderate currency depreciation through tighter controls on financial capital outflows. However, this route also comes with costs. EMDEs that tighten nonresident outflows will face reputational damage that would worsen their future access to international capital markets (Clayton and others 2022). Prohibitions exclusively targeting resident outflows might yield limited benefits while inflicting considerable domestic administrative and political costs. Supportive fiscal responses are largely off the table owing to higher sovereign debt levels.

The modern floating exchange rate system emerged fifty years ago amid conditions superficially much like today's: high inflation pressures, severe commodity price shocks, geopolitical tensions, and an inward turn by the United States from perceived burdens of global leadership. Inflation persisted in AEs until the early 1980s. But global disinflation, led by a strong dollar, threw many developing countries into a prolonged debt crisis and nearly a decade of lost growth during the 1980s. The restoration of price stability in the United States, coupled with the growth of US and world capital markets and deepening global trade links, eventually solidified the US dollar's de facto position as the dominant global currency, notwithstanding the scrapping of the de jure Bretton Woods arrangements that had centered on the dollar. The dollar's primacy was boosted further by US sponsorship of worldwide economic integration and opening after the collapse of the Soviet empire.

Strong contractionary measures by the world's central banks, acting with the relative independence they achieved largely as a result of past unpleasant inflation experiences, are likely to tame inflation this time. Indeed, there is a danger that central banks jointly create an unnecessarily sharp global recession through uncoordinated policies that effectively export inflation to trading partners through actions that strengthen their own currencies, as modeled by Oudiz and Sachs (1984) in this journal. In the present environment, central bankers need to be even more than usually attentive to the actions and reactions of their counterparts abroad.

The US macroeconomic outlook is once again central. Were it to remain unchecked, persistently high inflation in the United States could undermine the dollar's key global status as the inflation of the 1970s threatened to do. That would only add to a current trend toward global market fragmentation powered by nationalist political movements and international tensions. All countries would suffer.

As in the early 1970s, the reliability of US support for multilateralism in international relations will be crucial in determining the dollar's future. Reinforced by the United States' still dominant economic and geopolitical position, the substantial positive network externalities from worldwide dollar use mean that competitors such as the euro and yuan are unlikely to dislodge the dollar in the near term. Despite China's global ambitions for its currency, this is especially true for the yuan as long as China's financial markets remain relatively closed to foreign investors. But the case for the yuan becomes more plausible as China's economy grows relative to global output and as it gradually pursues targeted financial opening.[56] Sharper

56. On China's financial opening strategy and the prospects for the yuan as a global currency, see Clayton and others (2022) and Gourinchas (2021). Arslanalp, Eichengreen, and Simpson-Bell (2022) examine the much discussed recent decline of the dollar in international reserves (from more than 70 percent in 1999 to 59 percent in the last quarter of 2021) and show that only about a fourth of the decline reflects higher yuan holdings, the rest being diversification by reserve managers into nontraditional currencies.

political tensions between country blocs punctuated by further weaponization of trade and financial relations would accelerate the process. A world with multiple key currencies and the factors that bring it about could well change the positions of EMDEs in global markets and the policy regimes they adopt in response.

Going forward, global shocks associated with health emergencies, extreme weather, and cyber security breaches will likely add to the strains on world financial markets. Today's vast and interconnected dollar-centric world capital market looks strikingly different from its shape fifty years ago, but it may look very different still fifty years hence.

ACKNOWLEDGMENTS We are grateful to Fernando Eguren Martin, Zhengyang Jiang, Gian Maria Milesi-Ferretti, Emi Nakamura, Mikkel Plagborg-Møller, Hyun Song Shin, Jón Steinsson, and Mark Watson for helpful discussion; to Victoria de Quadros for outstanding research assistance; and to Egemen Eren, Gordon Liao, Friederike Niepmann, and Jianlin Wang for help obtaining data. Janice Eberly and our discussants, Şebnem Kalemli-Özcan and Matteo Maggiori, provided helpful detailed comments. We also had the benefit of reactions from *BPEA* conference participants. Obstfeld acknowledges with thanks financial support from the Class of 1958 chair and the Clausen Center at UC Berkeley. Zhou acknowledges generous financial support from the International Economics Section and the Griswold Center for Economic Policy Studies at Princeton University.

References

Adrian, Tobias, Richard K. Crump, and Emanuel Moench. 2013. "Pricing the Term Structure with Linear Regressions." *Journal of Financial Economics* 110, no. 1: 110–38.

Adrian, Tobias, and Hyun Song Shin. 2014. "Procyclical Leverage and Value-at-Risk." *Review of Financial Studies* 27, no. 2: 373–403.

Ammer, John, Michiel De Pooter, Christopher Erceg, and Steve Kamin. 2016. "International Spillovers of Monetary Policy." IFDP Notes. Washington: Board of Governors of the Federal Reserve System.

Arslanalp, Serkan, Barry J. Eichengreen, and Chima Simpson-Bell. 2022. "The Stealth Erosion of Dollar Dominance and the Rise of Nontraditional Reserve Currencies." *Journal of International Economics* 138:103656.

Avdjiev, Stefan, Valentina Bruno, Catherine Koch, and Hyun Song Shin. 2019. "The Dollar Exchange Rate as a Global Risk Factor: Evidence from Investment." *IMF Economic Review* 67, no. 1: 151–73.

Avdjiev, Stefan, Wenxin Du, Catherine Koch, and Hyun Song Shin. 2019. "The Dollar, Bank Leverage, and Deviations from Covered Interest Parity." *American Economic Review: Insights* 1, no. 2: 193–208.

Bahaj, Saleem, and Ricardo Reis. 2022. "Central Bank Swap Lines: Evidence on the Effects of the Lender of Last Resort." *Review of Economic Studies* 89, no. 4: 1654–93.

Bems, Rudolfs, Francesca Caselli, Francesco Grigoli, and Bertrand Gruss. 2021. "Expectations' Anchoring and Inflation Persistence." *Journal of International Economics* 132:103516.

Ben Zeev, Nadav. 2019. "Global Credit Supply Shocks and Exchange Rate Regimes." *Journal of International Economics* 116:1–32.

Bénétrix, Agustin, Deepali Gautam, Luciana Juvenal, and Martin Schmitz. 2019. "Cross-Border Currency Exposures." Working Paper 19/299. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2019/12/27/Cross-Border-Currency-Exposures-48876.

Bertaut, Carol, Bastian von Beschwitz, and Stephanie Curcuru. 2021. "The International Role of the U.S. Dollar." FEDS Notes. Washington: Board of Governors of the Federal Reserve System.

Bianchi, Javier, Saki Bigio, and Charles Engel. 2021. "Scrambling for Dollars: International Liquidity, Banks and Exchange Rates." Working Paper 29457. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29457.

Borio, Claudio, Robert N. McCauley, Patrick McGuire, and Vladyslav Sushko. 2016. "Covered Interest Parity Lost: Understanding the Cross-Currency Basis." *BIS Quarterly Review*, 45–64.

Boz, Emine, Camila Casas, Georgios Georgiadis, Gita Gopinath, Helena Le Mezo, Arnaud Mehl, and Tra Nguyen. 2022. "Patterns of Invoicing Currency in Global Trade: New Evidence." *Journal of International Economics* 136:103604.

Branson, William H. 1970. "Monetary Policy and the New View of International Capital Movements." *Brookings Papers on Economic Activity*, no. 2: 235–62.

Brave, Scott A., and David Kelley. 2017. "Introducing the Chicago Fed's New Adjusted National Financial Conditions Index." *Chicago Fed Letter*, no. 386. https://www.chicagofed.org/~/media/publications/chicago-fed-letter/2017/cfl386-pdf.pdf.

Bruno, Valentina, Se-Jik Kim, and Hyun Song Shin. 2018. "Exchange Rates and the Working Capital Channel of Trade Fluctuations." *American Economic Association Papers and Proceedings* 108:531–36.

Bruno, Valentina, and Hyun Song Shin. 2015a. "Capital Flows and the Risk-Taking Channel of Monetary Policy." *Journal of Monetary Economics* 71:119–32.

Bruno, Valentina, and Hyun Song Shin. 2015b. "Cross-Border Banking and Global Liquidity." *Review of Economic Studies* 82, no. 2: 535–64.

Bruno, Valentina, and Hyun Song Shin. 2021. "Dollar and Exports." Working Paper 819. Basel: Bank for International Settlements. https://www.bis.org/publ/work819.pdf.

Calvo, Guillermo A., Leonardo Leiderman, and Carmen M. Reinhart. 1996. "Inflows of Capital to Developing Countries in the 1990s." *Journal of Economic Perspectives* 10, no. 2: 123–39.

Canzoneri, Matthew, Robert Cumby, Behzad Diba, and David López-Salido. 2008. "Monetary Aggregates and Liquidity in a Neo-Wicksellian Framework." *Journal of Money, Credit and Banking* 40, no. 8: 1667–98.

Canzoneri, Matthew, Robert Cumby, Behzad Diba, and David López-Salido. 2013. "Key Currency Status: An Exorbitant Privilege and an Extraordinary Risk." *Journal of International Money and Finance* 37:371–93.

Carstens, Agustin, and Hyun Song Shin. 2019. "Emerging Markets Aren't Out of the Woods Yet." *Foreign Affairs*, March 15. https://www.foreignaffairs.com/world/emerging-markets-arent-out-woods-yet.

Cerutti, Eugenio M., Maurice Obstfeld, and Haonan Zhou. 2021. "Covered Interest Parity Deviations: Macrofinancial Determinants." *Journal of International Economics* 130:103447.

Cesa-Bianchi, Ambrogio, and Andrej Sokol. 2022. "Financial Shocks, Credit Spreads, and the International Credit Channel." *Journal of International Economics* 135:103543.

Chahrour, Ryan, and Rosen Valchev. 2022. "Trade Finance and the Durability of the Dollar." *Review of Economic Studies* 89, no. 4: 1873–910. https://academic.oup.com/restud/article-abstract/89/4/1873/6414211?redirectedFrom=fulltext.

Chinn, Menzie D., and Hiro Ito. 2006. "What Matters for Financial Development? Capital Controls, Institutions, and Interactions." *Journal of Development Economics* 81, no. 1: 163–92.

Clayton, Christopher, Amanda Dos Santos, Matteo Maggiori, and Jesse Schreger. 2022. "Internationalizing Like China." Working Paper 30336. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30336.

Committee on the Global Financial System. 2020. “US Dollar Funding: An International Perspective.” CGFS Paper 65. Basel: Bank for International Settlements.

Committee on the Global Financial System. 2021. “Changing Patterns of Capital Flows.” CGFS Paper 66. Basel: Bank for International Settlements.

Dahlquist, Magnus, and Paul Söderlind. 2022. “Individual Forecasts of Exchange Rates.” Research Paper 22-06. Stockholm: Swedish House of Finance.

Davis, J. Scott, Giorgio Valente, and Eric van Wincoop. 2021. “Global Drivers of Gross and Net Capital Flows.” *Journal of International Economics* 128:103397.

Dedola, Luca, Giorgios Georgiadis, Johannes Gräb, and Arnaud Mehl. 2021. “Does a Big Bazooka Matter? Quantitative Easing Policies and Exchange Rates.” *Journal of Monetary Economics* 117:489–506.

Degasperi, Riccardo, Seokki Simon Hong, and Giovanni Ricco. 2021. “The Global Transmission of U.S. Monetary Policy.” Working Paper. https://www.aeaweb.org/conference/2022/preliminary/paper/irQaY589.

De Leo, Pierre, Gita Gopinath, and Şebnem Kalemli-Özcan. 2022. “Monetary Policy Cyclicality in Emerging Economies.” Working Paper 30458. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30458.

Del Negro, Marci, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti. 2017. “Safety, Liquidity, and the Natural Rate of Interest.” *Brookings Papers on Economic Activity*, Spring, 235–94.

Dornbusch, Rudiger. 1976. “Expectations and Exchange Rate Dynamics.” *Journal of Political Economy* 84, no. 6: 1161–76.

Dornbusch, Rudiger. 1980. “Exchange Rate Economics: Where Do We Stand?” *Brookings Papers on Economic Activity*, no. 1: 143–85.

Driscoll, John C., and Aart C. Kraay. 1998. “Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data.” *Review of Economics and Statistics* 80, no. 4: 549–60.

Druck, Pablo, Nicholas E. Magud, and Rodrigo Mariscal. 2018. “Collateral Damage: Dollar Strength and Emerging ‘Markets’ Growth.” *North American Journal of Economics and Finance* 43:97–117.

Du, Wenxin. 2019. Commentary on “Mind the Gap in Sovereign Debt Markets: The U.S. Treasury Basis and the Dollar Risk Factor,” by Arvind Krishnamurthy and Hanno Lustig. In *Economic Policy Symposium Proceedings: Challenges for Monetary Policy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas.

Du, Wenxin, Joanne Im, and Jesse Schreger. 2018. “The U.S. Treasury Premium.” *Journal of International Economics* 112:167–81.

Du, Wenxin, Carolin E. Pflueger, and Jesse Schreger. 2020. “Sovereign Debt Portfolios, Bond Risks, and the Credibility of Monetary Policy.” *Journal of Finance* 75, no. 6: 3097–138.

Du, Wenxin, and Jesse Schreger. 2022. “CIP Deviations, the Dollar, and Frictions in International Capital Markets.” In *Handbook of International Economics, Volume 6*, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Amsterdam: North Holland.

Du, Wenxin, Alexander Tepper, and Adrien Verdelhan. 2018. “Deviations from Covered Interest Rate Parity.” *Journal of Finance* 73, no. 3: 915–57.

Eguren Martin, Fernando, Mayukh Mukhopadhyay, and Carlos van Hombeeck. 2017. “The Global Role of the US Dollar and Its Consequences.” *Bank of England Quarterly Bulletin* Q4.

Eichenbaum, Martin, Benjamin K. Johannsen, and Sergio Rebelo. 2021. “Monetary Policy and the Predictability of Nominal Exchange Rates.” *Review of Economic Studies* 88, no. 1: 192–228.

Eichengreen, Barry, and Poonam Gupta. 2018. “Managing Sudden Stops.” In *Monetary Policy and Global Spillovers: Mechanisms, Effects, and Policy Measures*, edited by Enrique G. Mendoza, Ernesto Pastén, and Diego Saravia. Santiago: Central Bank of Chile.

Engel, Charles. 2016. “Exchange Rates, Interest Rates, and the Risk Premium.” *American Economic Review* 106, no. 2: 436–74.

Engel, Charles, and Steve Pak Yeung Wu. Forthcoming. “Liquidity and Exchange Rates: An Empirical Investigation.” *Review of Economic Studies*.

Financial Stability Board. 2022. *US Dollar Funding and Emerging Market Economy Vulnerabilities*. https://www.fsb.org/2022/04/us-dollar-funding-and-emerging-market-economy-vulnerabilities/.

Forbes, Kristin J., and Francis E. Warnock. 2012. “Capital Flow Waves: Surges, Stops, Flight, and Retrenchment.” *Journal of International Economics* 88, no. 2: 235–51.

Gabaix, Xavier, and Matteo Maggiori. 2015. “International Liquidity and Exchange Rate Dynamics.” *Quarterly Journal of Economics* 130, no. 3: 1369–420.

Georgiadis, Georgios, Gernot J. Müller, and Ben Schumann. 2021. “Global Risk and the Dollar.” Discussion Paper 16245. London: Centre for Economic Policy Research.

Gertler, Mark, and Peter Karadi. 2015. “Monetary Policy Surprises, Credit Costs, and Economic Activity.” *American Economic Journal: Macroeconomics* 7, no. 1: 44–76.

Gilchrist, Simon, Bin Wei, Vivian Z. Yue, and Egon Zakrajšek. 2022. “Sovereign Risk and Financial Risk.” *Journal of International Economics* 136:103603.

Gilchrist, Simon, and Egon Zakrajšek. 2012. “Credit Spreads and Business Cycle Fluctuations.” *American Economic Review* 102, no. 4: 1692–720.

Goldberg, Linda S., and Fabiola Ravazzolo. 2022. “The Fed’s International Dollar Liquidity Facilities: New Evidence on Effects.” Working Paper 29982. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29982.

Gonçalves, Sílvia, Ana Maria Herrera, Lutz Kilian, and Elena Pesavento. 2022. “When Do State-Dependent Local Projections Work?” Working Paper 2205. Federal Reserve Bank of Dallas. https://www.dallasfed.org/research/papers/2022/wp2205.

Gopinath, Gita, and Jeremy C. Stein. 2021. “Banking, Trade, and the Making of a Dominant Currency.” *Quarterly Journal of Economics* 136, no. 2: 783–830.

Gorodnichenko, Yuriy, and Byoungchan Lee. 2020. “Forecast Error Variance Decompositions with Local Projections.” *Journal of Business and Economic Statistics* 38:921–33.

Gourinchas, Pierre-Olivier. 2018. “Monetary Policy Transmission in Emerging Markets: An Application to Chile.” In *Monetary Policy and Global Spillovers: Mechanisms, Effects, and Policy Measures*, edited by Enrique G. Mendoza, Ernesto Pastén, and Diego Saravia. Santiago: Central Bank of Chile.

Gourinchas, Pierre-Olivier. 2021. “The Dollar Hegemon? Evidence and Implications for Policymakers.” In *The Asian Monetary Policy Forum: Insights for Central Banking*, edited by Steven J. Davis, Edward S. Robinson, and Bernard Yeung. Singapore: World Scientific Publishing.

Gourinchas, Pierre-Olivier, Walker D. Ray, and Dimitri Vayanos. 2022. “A Preferred-Habitat Model of Term Premia, Exchange Rates, and Monetary Policy Spillovers.” Working Paper 29875. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29875.

Greenwood, Robin, Samuel G. Hanson, Jeremy C. Stein, and Adi Sunderam. 2020. “A Quantity-Driven Theory of Term Premia and Exchange Rates.” Working Paper 27615. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27615.

Ha, Jongrim, M. Ayhan Kose, and Franziska Ohnsorge. 2021. “One-Stop Source: A Global Database of Inflation.” Working Paper 9737. Washington: World Bank. https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-9737.

Hansen, Lars Peter, and Robert J. Hodrick. 1980. “Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis.” *Journal of Political Economy* 88, no. 5: 829–53.

Hofmann, Boris, and Taejin Park. 2020. “The Broad Dollar Exchange Rate as an EME Risk Factor.” *BIS Quarterly Review*, December.

Horn, Sebastian, Carmen M. Reinhart, and Christoph Trebesch. 2019. “China’s Overseas Lending.” Working Paper 26050. Cambridge, Mass.: National Bureau of Economic Research.

Ilzetzki, Ethan, Carmen M. Reinhart, and Kenneth S. Rogoff. 2019. “Exchange Arrangements Entering the Twenty-First Century: Which Anchor Will Hold?” *Quarterly Journal of Economics* 134, no. 2: 599–646.

International Monetary Fund (IMF). 2016. “Global Trade: What’s Behind the Slowdown?” In *World Economic Outlook: Subdued Demand: Symptoms and Remedies*. Washington: Author.

Itskhoki, Oleg. 2021. “The Story of the Real Exchange Rate.” *Annual Review of Economics* 13, no. 1: 423–55.

Itskhoki, Oleg, and Dmitry Mukhin. 2021. “Exchange Rate Disconnect in General Equilibrium.” *Journal of Political Economy* 129, no. 8: 2183–232.

Jiang, Zhengyang, Arvind Krishnamurthy, and Hanno Lustig. 2020. “Dollar Safety and the Global Financial Cycle.” Working Paper 27682. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27682.

Jiang, Zhengyang, Arvind Krishnamurthy, and Hanno Lustig. 2021. "Foreign Safe Asset Demand and the Dollar Exchange Rate." *Journal of Finance* 76, no. 3: 1049–89.

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95, no. 1: 161–82.

Jordà, Òscar, Moritz Schularick, Alan M. Taylor, and Felix Ward. 2019. "Global Financial Cycles and Risk Premiums." *IMF Economic Review* 67, no. 1: 109–50.

Kalemli-Özcan, Şebnem. 2019. "U.S. Monetary Policy and International Risk Spillovers." In *Economic Policy Symposium Proceedings: Challenges for Monetary Policy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas.

Kalemli-Özcan, Şebnem, and Liliana Varela. 2021. "Five Facts about the UIP Premium." Working Paper 28923. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28923.

Kekre, Rohan, and Moritz Lenel. 2021. "The Flight to Safety and International Risk Sharing." Working Paper 29238. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29238.

Klein, Michael W., and Jay C. Shambaugh. 2015. "Rounding the Corners of the Policy Trilemma: Sources of Monetary Policy Autonomy." *American Economic Journal: Macroeconomics* 7, no. 4: 33–66.

Kodres, Laura, Leslie Sheng Shen, and Darrell Duffie. 2022. "Dollar Funding Stresses in China." Working Paper. Social Science Research Network, May 25. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4099944.

Kouri, Pentti J. K., and Jorge Braga de Macedo. 1978. "Exchange Rates and the International Adjustment Process." *Brookings Papers on Economic Activity*, no. 1: 111–57.

Krishnamurthy, Arvind, and Hanno Lustig. 2019. "Mind the Gap in Sovereign Debt Markets: The U.S. Treasury Basis and the Dollar Risk Factor." In *Economic Policy Symposium Proceedings: Challenges for Monetary Policy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas.

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

Lane, Philip R., and Gian Maria Milesi-Ferretti. 2018. "The External Wealth of Nations Revisited: International Financial Integration in the Aftermath of the Global Financial Crisis." *IMF Economic Review* 66, no. 1: 189–222.

Liao, Gordon, and Tony Zhang. 2020. "The Hedging Channel of Exchange Rate Determination." Discussion Paper 1283. Washington: Board of Governors of the Federal Reserve System.

Lilley, Andrew, Matteo Maggiori, Brent Neiman, and Jesse Schreger. 2022. "Exchange Rate Reconnect." *Review of Economics and Statistics* 104, no. 4: 845–55.

Linnemann, Ludger, and Andreas Schabert. 2015. "Liquidity Premia and Interest Rate Parity." *Journal of International Economics* 97, no. 1: 178–92.

Liu, Zheng, Mark M. Spiegel, and Andrew Tai. 2017. "Measuring the Effects of Dollar Appreciation on Asia: A FAVAR Approach." *Journal of International Money and Finance* 74:353–70.

Loipersberger, Florian, and Johannes Matschke. 2022. "Financial Cycles and Domestic Policy Choices." *European Economic Review* 143:104034.

Maggiori, Matteo. 2022. "International Macroeconomics with Imperfect Financial Markets." In *Handbook of International Economics, Volume 6*, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Amsterdam: North Holland.

Maggiori, Matteo, Brent Neiman, and Jesse Schreger. 2020. "International Currencies and Capital Allocation." *Journal of Political Economy* 128, no. 6: 2019–66.

Miranda-Agrippino, Silvia, Tsvetelina Nenova, and Hélène Rey. 2020. "Global Footprints of Monetary Policies." Working Paper. http://www.helenerey.eu/AjaxRequestHandler.ashx?Function=GetSecuredDOC&DOCUrl=App_Data/helenerey_eu/Working-Papers_en-GB/_Documents_2019-20/291828257_67186463733_MirandaAgrippinoNenovaRey.pdf.

Miranda-Agrippino, Silvia, and Hélène Rey. 2020. "U.S. Monetary Policy and the Global Financial Cycle." *Review of Economic Studies* 87, no. 6: 2754–76.

Miranda-Agrippino, Silvia, and Hélène Rey. 2022. "The Global Financial Cycle." In *Handbook of International Economics, Volume 6*, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Amsterdam: North Holland.

Montiel Olea, José Luis, and Mikkel Plagborg-Møller. 2021. "Local Projection Inference Is Simpler and More Robust Than You Think." *Econometrica* 89, no. 4: 1789–823.

Mukhin, Dmitry. 2022. "An Equilibrium Model of the International Price System." *American Economic Review* 112, no. 2: 650–88.

Nagel, Stefan. 2016. "The Liquidity Premium of Near-Money Assets." *Quarterly Journal of Economics* 131, no. 4: 1927–71.

Obstfeld, Maurice. 2015. "Trilemmas and Trade-offs: Living with Financial Globalisation." Working Paper 480. Basel: Bank for International Settlements. https://www.bis.org/publ/work480.pdf.

Obstfeld, Maurice. 2021. "The Global Capital Market Reconsidered." *Oxford Review of Economic Policy* 37, no. 4: 690–706.

Obstfeld, Maurice. 2022. "The International Financial System after Covid-19." In *Prospects of the Global Economy after Covid-19*, edited by Thorsten Beck and Yung Chul Park. London: CEPR Press.

Obstfeld, Maurice, Jonathan D. Ostry, and Mahvash S. Qureshi. 2019. "A Tie That Binds: Revisiting the Trilemma in Emerging Market Economies." *Review of Economics and Statistics* 101, no. 2: 279–93.

Obstfeld, Maurice, and Alan M. Taylor. 2017. "International Monetary Relations: Taking Finance Seriously." *Journal of Economic Perspectives* 31, no. 3: 3–28.

Oudiz, Gilles, and Jeffrey Sachs. 1984. "Macroeconomic Policy Coordination among the Industrial Economies." *Brookings Papers on Economic Activity*, no. 1: 1–75.

Plagborg-Møller, Mikkel, and Christian K. Wolf. 2021. “Local Projections and VARs Estimate the Same Impulse Responses.” *Econometrica* 89, no. 2: 955–80.

Ramey, Valerie A., and Sarah Zubairy. 2018. “Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data.” *Journal of Political Economy* 126, no. 2: 850–901.

Reinhart, Carmen M., and Vincent R. Reinhart. 2009. “Capital Flow Bonanzas: An Encompassing View of the Past and Present.” In *NBER International Seminar on Macroeconomics 2008*, edited by Jeffrey Frankel and Christopher Pissarides. Chicago: University of Chicago Press.

Rey, Hélène. 2013. “Dilemma not Trilemma: The Global Financial Cycle and Monetary Policy Independence.” In *Economic Policy Symposium Proceedings: Global Dimensions of Unconventional Monetary Policy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas.

Securities Industry and Financial Markets Association (SIFMA). 2022. *SIFMA Research Quarterly—4Q21*. https://www.sifma.org/wp-content/uploads/2022/01/US-Research-Quarterly-Equity-2022-01-26-SIFMA.pdf.

Shin, Hyun Song. 2020. “Global Liquidity and Procyclicality.” In *The State of Economics, the State of the World*, edited by Kaushik Basu, David Rosenblatt, and Claudia Sepúlveda. Cambridge, Mass.: MIT Press.

Shousha, Samer. 2022. “The Dollar and Emerging Market Economies: Channels and Impacts.” Working Paper. https://www.dropbox.com/s/wc8vwjsngp5sj0e/Shousha_USDEME_Last.pdf?dl=0.

Stock, James H., and Mark W. Watson. 2018. “Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments.” *Economic Journal* 128, no. 610: 917–48.

Valchev, Rosen. 2020. “Bond Convenience Yields and Exchange Rate Dynamics.” *American Economic Journal: Macroeconomics* 12, no. 2: 124–66.

von Beschwitz, Bastian, Christopher G. Collins, and Deepa D. Datta. 2019. “Revisions to the Federal Reserve Dollar Indexes.” FEDS Notes. Washington: Board of Governors of the Federal Reserve System.

Wu, Jing Cynthia, and Fan Dora Xia. 2016. “Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound.” *Journal of Money, Credit and Banking* 48, no. 2-3: 253–91.

Comments and Discussion

COMMENT BY

ŞEBNEM KALEMLI-ÖZCAN A central issue in international macroeconomics regards the transmission of shocks between countries. As laid bare by the 2008 global financial crisis, the COVID-19 global pandemic, and the recent monetary policy tightening of central banks around the world, it is getting extremely difficult for policymakers to pursue domestic stabilization mandates in an increasingly interconnected global economy. Since a combination of financial and trade linkages ties domestic outcomes to global shocks, policies driven by domestic mandates will also have international spillover effects.

In this world, US monetary policy developments retain a major influence. A large body of literature shows that fluctuations in US monetary policy affect global investors' risk sentiments and, in turn, global financial conditions. The link between US monetary policy and global financial conditions forms the global financial cycle (GFC), as originally shown by Rey (2013). GFC is defined as the co-movement of risky asset prices, capital flows, financial intermediary leverage, and global growth. One common factor summarizing GFC is investors' risk sentiments measured by the CBOE Volatility Index (VIX), capturing global risk aversion and uncertainty.

The paper by Obstfeld and Zhou documents the same cycle, but instead of having it driven by US monetary policy or a measure of global risk aversion such as the VIX, they argue that the US dollar is in the driver's seat. They call it the global dollar cycle. As all these variables are endogenous and correlated with each other, the global dollar cycle is also correlated with the risk sentiments of global investors and with US monetary policy. The

Brookings Papers on Economic Activity, Fall 2022: 428–447

approach by Obstfeld and Zhou has a fundamental advantage in terms of documenting the quantitative importance of the financial channel of international spillovers over the standard trade channel. Since their key variable for the global dollar cycle is the US dollar's nominal exchange rate, they can account for tighter global financial conditions linked to a strong dollar simultaneously with the competitiveness of other countries' currencies that are depreciating against the US dollar when the cycle turns. As countries suffer from the global dollar cycle when the US dollar appreciates, experiencing slower growth in spite of higher net exports, it is clear that the financial channel dominates over the trade channel in the data.[1]

This is a very nice and timely contribution to *Brookings Papers on Economic Activity*. Showing the quantitative importance of the financial channel for international transmission of shocks and policies over the trade channel is especially important in a world where the share in trade and world output for emerging markets and developing economies (EMDEs) is larger than that of the United States, as documented by the authors. The trade channel tells us that if the US dollar is strong, EMDE currencies are weak, and this is good for their net exports, improving their current accounts.[2] By highlighting the quantitative importance of the financial channel, the paper sheds light on the perverse fact that EMDEs do worse when their currencies depreciate against the dollar. This fact has been documented by extensive literature focusing on contractionary depreciations in EMDEs. However, in this literature authors have had a hard time differentiating between the shocks driving the currency depreciations vis-à-vis the US dollar in EMDEs, since currency depreciations and financial crises go hand in hand: EMDE currencies tank at the same time those countries experience banking and sovereign crises. The "dollar shock" of the current paper is also not exogenous in a pure sense. Since the authors define the dollar shock to be an appreciation of the US dollar against advanced economies' (AEs)—Group of Ten (G10)—currencies, it can at least be taken as an external shock to EMDEs. When the US dollar appreciates against AEs, it also appreciates against EMDE currencies via a global appreciation of the US dollar.

1. An early contribution highlighting the role of the US dollar in bilateral nominal exchange rates for negative financial spillovers for other countries instead of positive trade spillovers is Bruno and Shin (2015).

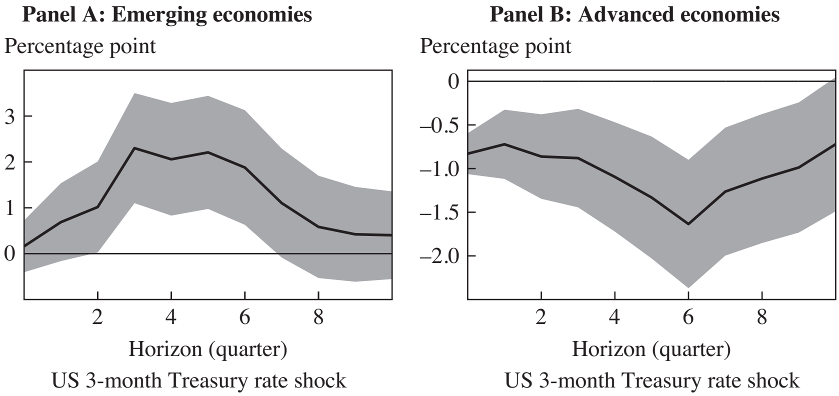
2. In a world where export prices are sticky in dollars, this channel might be muted; see Gopinath and others (2020).

The key result of the paper is how EMDEs got hurt from the dollar shock over quite a long horizon and from the global dollar cycle in general, whereas AEs, whose currencies are also depreciating against the US dollar, can get by. The authors document that dollar appreciation shocks predict declines in output, consumption, and investment, together with declines in domestic credit, terms of trade (higher import prices than export prices), and higher sovereign borrowing spreads on foreign currency debt in a sample of twenty-six EMDEs during 1999–2019. They also show that this result is more pronounced in EMDEs who peg their exchange rate, who did not adopt an inflation-targeting monetary policy framework, and who have high levels of external debt denominated in dollars.

The key policy questions then become: Why are EMDEs different? And what type of policies can EMDEs employ to deal with the global dollar cycle? Before getting into the answers given to these questions by the authors, let me briefly summarize what the standard international macro theory teaches us in terms of external shocks and EMDEs. For any small open economy, EMDE or not, standard theory postulates that countries should let their exchange rate carry the burden of adjustment when financial conditions change in the rest of the world. The intuition goes back to what I called the trade channel above. Monetary policy tightening slows down economic activity in the United States, which decreases US external demand. However, the associated appreciation of the dollar (depreciation in the rest of the world) helps other countries increase their exports to the United States and cut back their imports from the United States. If these countries are also net borrowers and experience capital outflows due to tightening of monetary policy in the United States, then a depreciating currency is the only force available to combat reduced activity by switching external demand to their goods. This channel, known as the expenditure switching channel of the Mundell-Fleming model, highlights the virtue of flexible exchange rates. It has been challenged in the academic literature and by the policymakers on the basis of the negative effects of excessive exchange rate volatility on countries with extensive debt denominated in the US dollars, dubbed as “fear of floating” by Calvo and Reinhart (2002). Hence, policies in support of limiting the exchange rate volatility have been used extensively by policymakers as foreign exchange interventions. There are also papers that show the optimality of such intervention policies depending on modeling of financial frictions (Itskhoki and Mukhin 2021).

If the main channel of international spillovers is the financial channel, the above reasoning changes. I argue (Kalemli-Özcan 2019) that exchange rate flexibility helps countries to smooth out the effects of financial cycles

Figure 1. The Effects of US Monetary Tightening on Borrowing Spreads in Emerging Markets and Advanced Economies



Source: Kalemli-Özcan (2019); reproduced with permission from Federal Reserve Bank of Kansas City.

driven by US monetary policy and a strong US dollar, such as the GFC and the global dollar cycle. My reasoning is not based on the standard expenditure switching channel but rather the strength of the primary channel of spillovers of these financial cycles in the data—changes in global investors' risk sentiments. I document that US monetary policy tightening leading to a strong US dollar increases the required excess return on EMDE bonds, which leads to contractionary outcomes in EMDEs, as also documented here by Obstfeld and Zhou. If EMDEs have flexible exchange rates, the risk premium increases are achieved in part through a currency depreciation. Under a pegged exchange rate, however, a sharper domestic monetary contraction would be needed to achieve the same risk premium rise with more damage to the economy. I show that free-floating EMDEs are much more insulated from risk premia shocks driven by US monetary policy than EMDEs with managed floats, in terms of their output.

Figure 1 is at the heart of my argument (Kalemli-Özcan 2019). The figure shows local projections, similar to Obstfeld and Zhou, but uses Gertler and Karadi's (2015) instrumented exogenous US monetary policy tightening between 1996 and 2018, instead of Obstfeld and Zhou's dollar shock.[3] As shown in panel A, EMDE government bond spreads vis-à-vis

3. A similar result is shown by Di Giovanni and others (2022) using the VIX instead of exogenous US monetary policy tightening.

US Treasuries, for less than twelve-month bonds, goes up more than one-to-one when the United States tightens, but for AEs, these spreads go down, as shown in panel B.

The reasoning is as follows. When the US monetary policy is tighter and the US dollar is stronger, global investors go to a risk-off mode and de-lever by shedding risk assets worldwide, not only domestically in the US stock markets. Since EMDEs are a riskier asset class as a whole than AEs, regardless of the type of investment (bonds, loans, equities) global investors make, EMDEs are affected much more, not only via de-leveraging but also via higher risk premia. The risk-averse financial intermediaries and the associated risk premia have a key role in this argument as the US financial intermediaries are pricing the risky assets worldwide. When the US dollar is stronger, global financial intermediaries get out of non-dollar assets and retrench to the US Treasuries. The fact that this happens by getting more out of EMDE assets (sudden stops) at the higher risk premia charged to EMDE assets highlights the key role of endogenously risk-averse global financial intermediaries as modeled by Akinci, Kalemli-Özcan, and Queralto (2022).[4] As a result, flexible exchange rates will help to smooth out this risk premia.

What other policy options are available to EMDEs, in addition to flexible exchange rates, to deal with the contractionary effects of the global dollar cycle and GFC? Using monetary policy to defend a currency against the US dollar can be counterproductive. As shown in Kalemli-Özcan (2019), EMDEs who try to prevent depreciations against a strong US dollar using monetary policy end up having larger contractions and higher risk premia. De Leo, Gopinath, and Kalemli-Özcan (2022) show that even if EMDEs run monetary expansions as a response to tight US policy, they still suffer the same contractionary outcomes. This result is consistent with the results of Obstfeld and Zhou as they show that floaters do not increase their monetary policy rates as a response to a strong US dollar but still suffer contractionary effects. These results open the door for other policies.[5]

In Kalemli-Özcan (2019) I state:

> Countries can act on the transmission channel cyclically by limiting credit growth and leverage during the booms and doing reverse during downturns. This can be achieved by the use of macroprudential policies. . . . The policies that limit

4. Gabaix and Maggiori (2015) and Itskhoki and Mukhin (2021) also rationalize these results, using a different theoretical mechanism than endogenous risk aversion. In those models, financial markets are segmented, limiting the amount of bonds on financial intermediaries' balance sheets.

5. The Integrated Policy Framework of the IMF models several optimal policies (Basu and others 2020).

> un-hedged foreign currency denominated liabilities not only in the financial sector but also in the nonfinancial corporate sector must be a priority. The rationale for these policies is to provide insulation from spillovers that arise from balance sheet effects of exchange rate fluctuations with large levels of un-hedged foreign currency denominated debt. . . .
>
> However, dealing with excessive credit growth and foreign currency denominated debt may not be enough. A significant component of international risk spillovers for EMEs is related to country-specific risk. . . . Long-run improvements in the quality and transparency of institutions will reduce idiosyncratic country risk and reduce the sensitivity of capital flows in EMEs to global risk premia and to foreign investors risk perceptions. . . . Strong institutions will also provide the needed credibility for implementing desirable macroprudential policies, to dampen the severe effects of financial cycles.
>
> . . . A collective reform agenda aimed at improving transparency, governance, accountability, fighting with corruption, protecting institutional integrity, and improving bureaucratic quality with an emphasis on central bank independence will be beneficial in terms of attracting long-term stable capital flows. These policies reduce the sensitivity of capital flows to changes in the center country monetary policy and associated risk sentiments. (156–58)

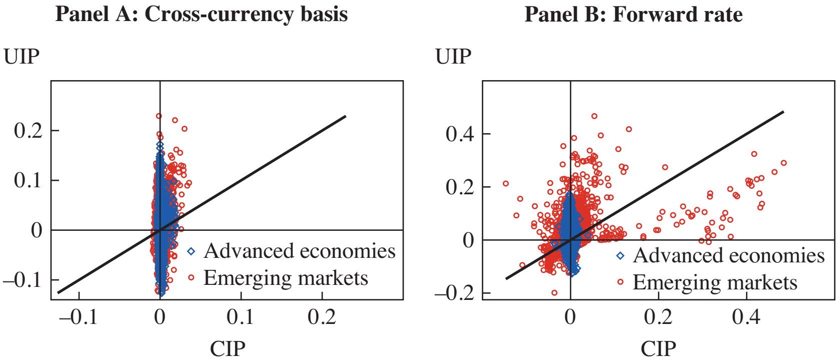
Interestingly, Obstfeld and Zhou reach the exact same policy conclusion: "more-flexible exchange rate regimes do not shut out the global financial cycle, but they are indeed helpful in buffering external financial shocks and can do so most effectively when supported by relatively high inflation credibility at the central bank and relatively low external dollarization."

These same policy conclusions in Kalemli-Özcan (2019) and in Obstfeld and Zhou suggest that the risk premia channel underlines the negative effects of both the GFC and the global dollar cycle on EMDEs. In Kalemli-Özcan (2019) I show that uncovered interest rate parity (UIP) deviations in EMDEs vis-à-vis the US dollar are a strong correlate of the risk premia channel. This is because UIP deviations are endogenous to both domestic and US monetary policy. There are several models in the literature that work through endogenous or exogenous UIP deviations. Obstfeld and Zhou provide a very useful framework to connect different pieces of this literature through the simple equation below.

UIP deviations, λ^{e}_{t+h}, can be written as the nonzero difference between the country and the US interest rate differentials minus the expected depreciation over h horizon in the nominal exchange rate defined as the local currency per the US dollar (so expected US dollar appreciation if the country interest rate is higher than that of the United States):

$$\lambda^{e}_{t+h} = \underbrace{\left(i_t - i^{US}_t\right)}_{IR\ Differential} - \underbrace{\left(s^{e}_{t+h} - s_t\right)}_{ER\ Adjustment} \neq 0 = \underbrace{\gamma^{\$}_t}_{convenience/liquidity\ premium} + \underbrace{\rho_t}_{excess\ returns}$$

Figure 2. Incomplete Pass-Through between UIP and CIP Deviations



Source: Kalemli-Özcan and Varela (2022); reproduced with authors' permission.

Using this equation, Obstfeld and Zhou equate the UIP deviations, where they formulate these deviations only for AE currencies vis-à-vis the US dollar, to a convenience/liquidity yield of the US Treasuries term plus an excess return term for AE currencies. They call the excess return part of the equation "dark matter." Risk aversion and financial frictions drive this part. One may consider excess returns as coming from risk-averse global intermediaries or financial frictions on global intermediary balance sheets or both. If we want to extend these UIP deviations to EMDEs, then adding a local part to the excess returns term is important as EMDE currencies provide much higher and persistent excess returns than AEs in the data (Kalemli-Özcan and Varela 2021). So dark matter is:

$$Dark\ Matter = \rho_t = \rho_t^{US} + \rho_t^{COUNTRY} = Global + Local$$

The authors focus on the relation between UIP/CIP (covered interest parity) deviations and the convenience/liquidity yield of US Treasuries. I want to make the point that these deviations look very different in the data, and there is no one-to-one pass-through between UIP and CIP deviations, as most often assumed in the literature. As shown in figure 2, regardless of measuring the CIP deviations with cross-currency basis swaps taken from Du and Schreger (2022) in panel A or CIP deviations based on forward rates in panel B, they are nowhere close to the 45 degree line, hence drastically different objects than UIP deviations. This suggests that understanding the underlying primitive behind the dollar shocks and why a strong dollar leads to contractionary responses in EMDEs but not in AEs requires an understanding of both UIP and CIP deviations. The convenience/liquidity yield of US Treasuries can be correlated with both, but due to very different reasons. CIP deviations are about arbitrage failing in hedging markets, where investors insure for currency risk, whereas UIP deviations can be about a higher price of risk for local currency bonds from which investors are expecting to earn higher excess returns.

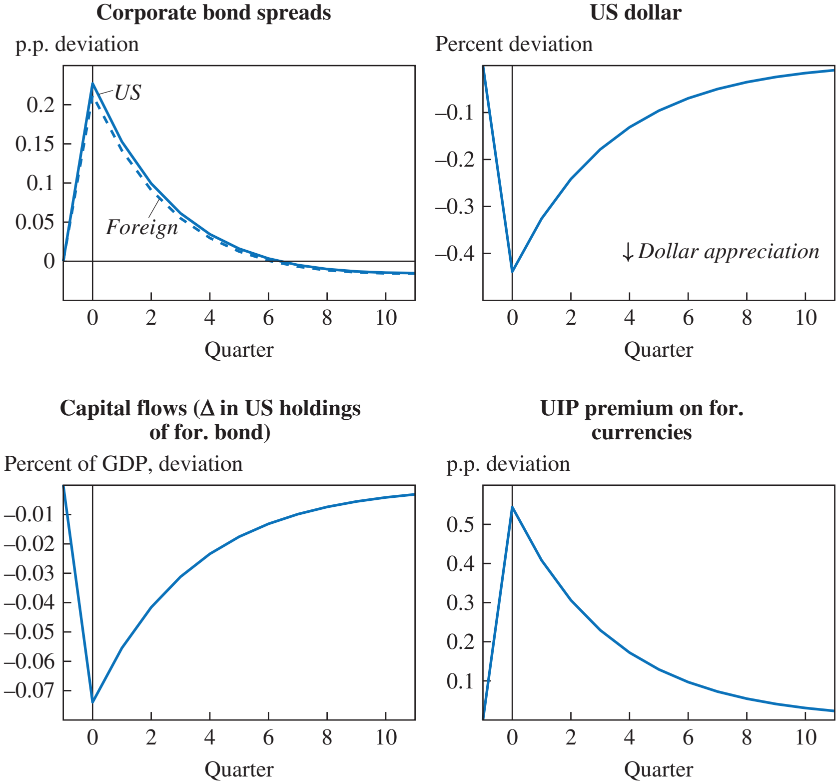
In fact, the authors' mechanism is exactly about these types of UIP deviations; when they dig deeper on what causes the dollar shock, they lean toward the explanation of the risk sentiments of global financial intermediaries. The million dollar question is how to measure the changing risk sentiments of global investors in the data. In the literature others have tried to do this using several different measures of investors' risk sentiments that are independent of the response of US monetary policy to developments in the US economy, such as the VIX or exogenous shocks to US monetary policy. The latter is estimated based on high-frequency identification capturing surprise reactions in financial markets to the Federal Reserve's decisions. The authors argue that these exogenous US monetary policy shocks can account for only a small share of dollar variability or sudden stops for EMDEs and that their dollar shock is broader. This is true, but the advantage of using exogenous tightening of US monetary policy is the ease of interpretation. The dollar shock is the dollar's weighted nominal exchange rate against other AEs, which is why it's not obvious how to interpret it structurally. What drives the strong US dollar? US monetary policy? Expansionary global economic activity? Contractionary financial conditions? The answer is important in order to be able to answer questions about EMDEs' optimal policy response and international coordination of monetary policies.

The authors are aware of this problem, which is why they try to instrument their dollar shock with a measure that picks up the risk sentiments of investors. They use the excess bond premium (EBP) of Gilchrist and Zakrajšek (2012). As EBP is the most significant correlate of CIP deviations, this measure fits well with the authors' purposes. But it is still not clear why this measure is a good measure to pick up financial intermediaries' risk sentiments. EBP is the residual spread in US corporate bonds after cleaning the default risk. An increase in EBP is shown in the literature to be correlated with de-leveraging of the financial sector and a slowdown in economic activity. In the literature it is interpreted as a measure capturing a reduction in the risk-bearing capacity of the financial sector as it relates to a contraction in the supply of credit in the United States.

Why would a higher EBP lead to contractionary outcomes in EMDEs? Going back to the authors' general framework, if the balance sheet constraint of global financial intermediaries prevents arbitrage, leading to CIP deviations only, this would not lead to higher spreads on local currency borrowing by EMDEs (UIP deviations), and hence it becomes harder to link real contractionary outcomes both to capital outflows and higher spreads. If, however, the existing balance sheet constraint of long-lived and forward-looking global financial intermediaries will be even tighter in the future due to higher pricing of currency risk, then all data facts can be explained, where higher risk sentiment is captured by EBP. Akinci, Kalemli-Özcan, and Queralto (2022) provide such a model. The reason why global financial intermediaries become more risk averse and want to de-lever their EMDE assets (captured by higher EBP premium) is an exogenous increase in uncertainty that can be captured by stock market volatility in the United States. Such earnings volatility in the United States hurts the balance sheets of the US intermediaries, and hence they become endogenously more risk averse and want to get out of their other risky investments such as EMDE assets. Such uncertainty spillovers can generate all the facts shown by Obstfeld and Zhou, as seen in figure 3: strong dollar, higher UIP deviations on EMDEs (higher emerging markets bond index and other spreads), sudden stops in EMDEs (capital outflows), depreciating EMDE currencies, and contractionary outcomes in EMDEs.

To conclude, this is a valuable paper providing a unifying framework on how to think about the global dollar cycle and the GFC and their detrimental effects, especially on EMDEs. The policy implication is clear: the case for flexible exchange rates is stronger. If the contractionary effects of a strong dollar work through higher excess bond premia, this means that risk spillovers of US monetary policy, as originally argued (Kalemli-Özcan 2019), is central to understanding the negative effects of a strong US dollar on EMDEs. In a world of risk spillovers, the coordination of monetary policy will be much more difficult; if any country loses its own monetary policy credibility for the sake of international coordination, risk premia can be higher due to higher uncertainty in financial markets, leading to worse contractionary outcomes in EMDEs (Coy 2022). This is why Obstfeld and Zhou conclude that flexible exchange rate regimes can buffer external financial shocks most effectively when supported by relatively high inflation credibility at the central banks.

Figure 3. The Effects of US Risk-Off Shocks in the Rest of the World



Source: Akinci, Kalemli-Özcan, and Queralto (2022); reproduced with authors' permission.

REFERENCES FOR THE KALEMLI-ÖZCAN COMMENT

Akinci, Ozge, Şebnem Kalemli-Özcan, and Albert Queralto. 2022. "Uncertainty Shocks, Capital Flows, and International Risk Spillovers." Working Paper 30026. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30026.

Basu, Suman S., Emine Boz, Gita Gopinath, Francisco Roch, and Filiz D. Unsal. 2020. "A Conceptual Model for the Integrated Policy Framework." Working Paper 2020/121. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2020/07/07/A-Conceptual-Model-for-the-Integrated-Policy-Framework-49558.

Bruno, Valentina, and Hyun Song Shin. 2015. "Capital Flows and the Risk-Taking Channel of Monetary Policy." *Journal of Monetary Economics* 71:119–32.

Calvo, Guillermo A., and Carmen M. Reinhart. 2002. "Fear of Floating." *Quarterly Journal of Economics* 107, no. 2: 379–408.

Coy, Peter. 2022. "A Strong Dollar Is Hurting Other Countries. Should the Fed Care?" *New York Times*, September 30. https://www.nytimes.com/2022/09/30/opinion/us-dollar-global-impact.html.

De Leo, Pierre, Gita Gopinath, and Şebnem Kalemli-Özcan. 2022. "Monetary Policy Cyclicality in Emerging Economies." Working Paper 30458. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w30458.

Di Giovanni, Julian, Şebnem Kalemli-Özcan, Mehmet Fatih Ulu, and Yusuf Soner Baskaya. 2022. "International Spillovers and Local Credit Cycles." *Review of Economic Studies* 89, no. 2:733–73.

Du, Wenxin, and Jesse Schreger. 2022. "CIP Deviations, the Dollar, and Frictions in International Capital Markets." In *Handbook of International Economics, Volume 6*, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Amsterdam: North Holland.

Gabaix, Xavier, and Matteo Maggiori. 2015. "International Liquidity and Exchange Rate Dynamics." *Quarterly Journal of Economics* 130, no. 3: 1369–420.

Gertler, Mark, and Peter Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7, no. 1: 44–76.

Gilchrist, Simon, and Egon Zakrajšek. 2012. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review* 102, no. 4: 1692–720.

Gopinath, Gita, Emine Boz, Camila Casas, Federico J. Díez, Pierre-Olivier Gourinchas, and Mikkel Plagborg-Møller. 2020. "Dominant Currency Paradigm." *American Economic Review* 110, no. 3: 677–719.

Itskhoki, Oleg, and Dmitry Mukhin. 2021. "Exchange Rate Disconnect in General Equilibrium." *Journal of Political Economy* 129, no. 8: 2183–232.

Kalemli-Özcan, Şebnem. 2019. "U.S. Monetary Policy and International Risk Spillovers." In *Economic Policy Symposium Proceedings: Challenges for Monetary Policy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas.

Kalemli-Özcan, Şebnem, and Liliana Varela. 2021. "Five Facts about the UIP Premium." Working Paper 28923. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28923.

Rey, Hélène. 2013. "Dilemma Not Trilemma: The Global Financial Cycle and Monetary Policy Independence." In *Economic Policy Symposium Proceedings: Global Dimensions of Unconventional Monetary Policy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas.

COMMENT BY

MATTEO MAGGIORI[1] It is a pleasure to have the opportunity to discuss this paper by Maurice Obstfeld and Haonan Zhou on the role of the dollar nominal exchange rate in transmitting financial conditions globally,

1. I am grateful to Janice Eberly and James Stock for inviting me to discuss this paper at the *BPEA* annual conference, and to Maurice Obstfeld and Haonan Zhou for interesting discussions of the contents of their paper. *BPEA* provided an honorarium for this discussion.

especially to emerging economies. The paper builds on a view of the determination of the dollar exchange rate in imperfect financial markets, whereby financial flows more than macro fundamentals pin down the exchange rate, and then goes on to provide evidence that a dollar appreciation negatively affects real economic activity in emerging economies. That is, dollar appreciation shocks, which tend to occur at times of global financial stress, predict economic downturns in emerging economies.

The paper is particularly timely, since in 2022 the dollar appreciated strongly, and many commentators and policymakers worry that soon we will observe the consequent deleterious effects on emerging economies. It is too soon to tell whether, in this instance, the view and evidence put forward by the authors will pan out. It is certainly not my comparative advantage to make such forecasts, and my reading of the empirical evidence to date leaves me wary of making such forecasts. I mention it here, instead, with the intent of highlighting the importance of the topic this paper is devoted to, and the need to assess what we know thus far and where academic research might go hunting for more evidence next.

In this respect, my summary judgment is that the paper is an excellent overview of the current stock of knowledge. It provides new evidence on the dollar's importance in shaping the global business cycle and highlights some challenges for policy in addressing unwanted swings in exchange rates. I recommend it as a read for both practitioners and academics.

THE IMPORTANCE OF GLOBAL GROSS POSITIONS Obstfeld and Zhou start their analysis by reminding the reader of the fast increase in global gross external positions that has occurred since the collapse of the Bretton Woods system in 1973. The liberalization of capital accounts which accompanied the move to a floating exchange rate system has been followed by an explosion in global financial positions. Figure 1 in the paper provides a great summary of the evidence. To clarify, a country's foreign asset is an asset that the country's domestic residents own abroad. Similarly, a country's foreign liability is an asset that residents of foreign countries own in that country. Figure 1 in the paper plots the sum of all countries' foreign assets and liabilities scaled by gross domestic product (GDP). These gross positions moved from being a small fraction of GDP in the early 1970s to a multiple of GDP today: a remarkable growth over this period.

These positions are interesting and consequential for several different reasons. The authors focus on two key elements. First, these large positions have to be intermediated at least in the medium run by financial intermediaries with limited risk-bearing capacity. The intermediaries' willingness to absorb the exchange rate risk is an important determinant of the level and dynamics of exchange rates. Second, many of these positions are in dollars, even when the United States is neither the holder nor the issuer. Further, financial (funding) conditions in the dollar market might affect the intermediaries' willingness to bear risks since much of the world financial sector is funded in dollars. I will briefly discuss each of these two elements; I tend to agree strongly on both points, even while recognizing that much progress is yet to be made in understanding these topics.

EXCHANGE RATE DETERMINATION WITH IMPERFECT FINANCIAL MARKETS The field of international macroeconomics and finance has in recent years progressed both empirically and theoretically by focusing on the question of who owns which assets around the world. On the theoretical front, this has required not only new models but also, in some cases, going back to older insights that had been largely forgotten, such as the portfolio balance theories in the 1970s. On the empirical front, we have witnessed, starting in 2007, the breakdown of the covered interest parity (CIP) condition, a central condition for the absence of arbitrage. It is rare to witness such a dramatic change in a basic condition of one of the most established financial markets. This prime evidence for currency market segmentation, reviewed extensively by Du and Schreger (2022), is inconsistent with models of perfect financial markets, including those that generate imperfect substitutability among currencies via risk premia.

The authors rightly focus on the recent literature that used financial frictions as a foundation for imperfect substitutability of assets in different currencies. The presence of market segmentation and financial frictions generates a set of specific predictions: CIP deviations, a direct effect of gross portfolio flows on exchange rates, and the effectiveness of foreign exchange intervention. It also casts a different light on classic stylized facts in the field, such as the disconnect of exchange rates from macro fundamentals and the carry trade.

In models with imperfect financial intermediation, the exchange rate is pinned down by imbalances in the demand and supply of assets in different currencies and, crucially, by the limited risk-bearing capacity of financiers that absorb these imbalances. The demand for the assets, the resulting gross capital flows, or the financiers' risk-bearing capacity might only have a distant relation with macro fundamentals, thus contributing to generating the disconnect. By placing global portfolios at center stage, this line of research stresses the importance of better data to understand these financial forces and their impact on the real economy, an ongoing effort in the field.

The intellectual origin of this modeling can be traced back to the Nurkse (1944) view of capital flows as inducing volatile and destabilizing exchange rate movements. The field has been inspired by the pioneering work of Pentti Kouri (1976, 1983). At the core of the portfolio balance approach is the idea of imperfect substitutability of assets denominated in different currencies. This contrasts with the traditional macroeconomics approach of imposing, either explicitly or implicitly via solution methods, the uncovered interest rate parity (UIP) condition of perfect substitutability. Gabaix and Maggiori (2015) provide a simple general equilibrium framework of the portfolio-balance determination of exchange rates under segmented currency markets, and Maggiori (2022) reviews the growing literature on this exchange rate determination framework.

THE DOLLAR AS AN INTERNATIONAL CURRENCY IN GLOBAL POSITIONS The authors rightly emphasize the dominance of the dollar in global positions and capital flows. Figure 7 in the paper provides an overview of the usage of the dollar to denominate cross-border debt (bonds and loans), settle payments (SWIFT), and more generally in foreign exchange transactions. In all these dimensions the dollar is used to a greater extent than what the economic size of the United States alone would predict. Plainly, the dollar is being used as an international currency, and used in relationships that never directly involve the United States. For example, think of a eurozone investor buying dollar denominated bonds of a Brazilian corporation.

The authors emphasize the importance of the centrality of the dollar in three main respects: (1) demand for dollar (safe) debt increases when global conditions worsen and, via the equilibrium determination described above, the dollar appreciates (Jiang, Krishnamurthy, and Lustig 2020); (2) intermediaries fund themselves in dollars, and when funding becomes tight in times of stress this makes the intermediaries less willing to absorb risk, including currency risk (Avdjiev and others 2019); and (3) the dollar importance in asset and goods trade denomination in emerging markets leaves these economies vulnerable to swings in the dollar exchange rate. The third aspect is crucial in transmitting variation in the dollar exchange rate to real economic activity in emerging economies.

THE DOLLAR AND EMERGING ECONOMIES' REAL PERFORMANCE The authors provide new evidence that dollar appreciation shocks predict downturns in emerging economies over the next few quarters. This is the most innovative part of the paper and the more relevant for policy. Figure 8 in the paper provides a great summary of the authors' main results. A 10 percent dollar appreciation against a basket of developed currencies is associated with a persistent fall in the GDP of emerging economies over the following eight quarters, reaching a trough of about −1.5 percent compared to trend. Investment also falls and the local currency depreciates.

The authors measure the dollar exchange rate against a basket of advanced economies' currencies to avoid any direct effect of local currencies of emerging economies contaminating their results. They also break down their analysis by the exchange rate and monetary policy regime and the amount of dollar external liabilities for each emerging economy. Overall, the analysis is careful, and the results are appropriately caveated.

There is much we do not know about how dollar exchange rate movements affect real activity in emerging markets. Which channels are important? Is the relationship stable or highly state dependent? How have changes in the policy response of emerging economies affected (or are likely to affect going forward) this relationship? The answers to these questions are beyond the scope of the current paper, and rightly so; solid answers will need future contributions to the literature, not a single paper. My own reading of the existing evidence is that it is supportive of a negative effect of dollar appreciation on emerging economies, but in judging the quality and extent of the evidence the proverbial bottle is at best half full. To fill up the bottle, future literature should focus on identification and carefully tracing each of the channels using micro data (and of course aggregation to macro). This is not just the call of an academic discussant for more evidence and careful work, it is a cautionary statement not to be overconfident in policy about these mechanisms given the state of the evidence.

WHAT CAN EMERGING MARKETS' POLICYMAKERS DO ABOUT THIS? The policy framework, especially in emerging markets, has evolved substantially since the global financial crisis of 2007–2009. Policies such as ex ante capital controls and foreign exchange interventions are now an integral part of the policy tool kit. These policies were previously regarded with diffidence, especially by Washington multilateral institutions. Such changes came about from the interplay of actual events (like crises); policy experimentations, often with emerging market central banks further ahead than the policy consensus; academic research on what inefficiencies these policies might help address and how; and finally multilateral discussions. I regard the changes that have taken place as beneficial, in the sense that welfare is probably higher with these policies in use in their current imperfect form than it would have been otherwise.

Especially in emerging markets these policies are used to tame waves of foreign capital that wash up on their shores. In buoyant times in global markets, capital chases assets in emerging economies in search for higher returns. When the eventual fall in global risk appetite comes, the fast withdrawal of capital produces welfare losses due to either the fire sale of local assets (a pecuniary externality) or a disproportionate fall in demand (a demand externality). The boom-and-bust pattern is excessive from an optimal policy perspective. Policy interventions aim to reduce the bust a lot by reducing the boom a little. Bianchi and Lorenzoni (2022) offer a great review of these issues, and Basu and others (2020) and Adrian and others (2022) provide a view into the IMF's evolving thinking about these policies.

An interesting question, based on the paper, is to what extent emerging market policymakers should aim to counteract the effect that movements in the dollar exchange rate might have on their economies. And if so, with which policies? Monetary policy alone might be limited or too blunt of a tool. In fact, it is not entirely obvious if the right response should be higher local interest rates in the hope of limiting local currency depreciation or lower rates to stimulate domestic demand. The answers might very well depend on the underlying channel through which the dollar is affecting the local economy.

Ex ante macroprudential policy in the form of capital controls or foreign exchange intervention to smooth exchange rate movements might be necessary to complement monetary policy and target specific margins like foreign capital flows or the exchange rate. These policies, however, are not free of problems. A prominent one is that these policies could be prone to abuse, especially in countries with a weak institutional framework. In this light, capital controls implemented as taxes are more likely to be abused as another way to inefficiently generate fiscal revenue. To minimize these risks, these policies should be directed by an independent body, like the central bank, with the best hope of isolating their implementation from political abuse.

CONCLUSION The paper both consolidates and pushes in new directions the existing literature on the role of the dollar in transmitting and setting global financial conditions. Swings in the value of the dollar are potentially affecting emerging economies, and policy has a role to play in reducing potential inefficiencies coming from boom-and-bust cycles. It is a timely paper worth reading and thinking about, as the world in 2022 is witnessing a strong appreciation of the dollar and many policymakers in emerging economies are contemplating how to react to this environment.

REFERENCES FOR THE MAGGIORI COMMENT

Adrian, Tobias, Christopher J. Erceg, Jesper Linde, Marcin Kolasa, and Pawel Zabczyk. 2022. "Managing Monetary Tradeoffs in Vulnerable Open Economies."

Discussion Paper 16972. London: Centre for Economic Policy Research. https://cepr.org/publications/dp16972.

Avdjiev, Stefan, Wenxin Du, Catherine Koch, and Hyun Song Shin. 2019. "The Dollar, Bank Leverage, and Deviations from Covered Interest Parity." *American Economic Review: Insights* 1, no. 2: 193–208.

Basu, Suman S., Emine Boz, Gita Gopinath, Francisco Roch, and Filiz D. Unsal. 2020. "A Conceptual Model for the Integrated Policy Framework." Working Paper 2020/121. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2020/07/07/A-Conceptual-Model-for-the-Integrated-Policy-Framework-49558.

Bianchi, J., and G. Lorenzoni. 2022. "The Prudential Use of Capital Controls and Foreign Currency Reserves." In *Handbook of International Economics, Volume 6*, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Amsterdam: North Holland.

Du, Wenxin, and Jesse Schreger. 2022. "CIP Deviations, the Dollar, and Frictions in International Capital Markets." In *Handbook of International Economics, Volume 6*, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Amsterdam: North Holland.

Gabaix, Xavier, and Matteo Maggiori. 2015. "International Liquidity and Exchange Rate Dynamics." *Quarterly Journal of Economics* 130, no. 3: 1369–420.

Jiang, Zhengyang, Arvind Krishnamurthy, and Hanno Lustig. 2020. "Dollar Safety and the Global Financial Cycle." Working Paper 27682. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27682.

Kouri, Pentti J. K. 1976. "The Exchange Rate and the Balance of Payments in the Short Run and in the Long Run: A Monetary Approach." *Scandinavian Journal of Economics* 78, no. 2: 280–304.

Kouri, Pentti J. K. 1983. "Balance of Payments and the Foreign Exchange Market: A Dynamic Partial Equilibrium Model." In *Economic Interdependence and Flexible Exchange Rates*, edited by Jagdeep S. Bhandari and Bluford Putnam. Cambridge, Mass.: MIT Press.

Maggiori, Matteo. 2022. "International Macroeconomics with Imperfect Financial Markets." In *Handbook of International Economics, Volume 6*, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Amsterdam: North Holland.

Nurkse, Ragnar. 1944. *International Currency Experience: Lessons of the Interwar Period*. Geneva: League of Nations.

GENERAL DISCUSSION Pierre-Olivier Gourinchas began the discussion by considering the current dollar appreciation episode and the absence of significant turmoil in emerging markets and developing economies (EMDEs) thus far. The United States is in the midst of sharp dollar appreciation, particularly in comparison to other advanced economies (AEs), Gourinchas pointed out. This paper, in addition to existing literature, demonstrates how tightening financial conditions associated with dollar appreciation indicate trouble for EMDEs.[1] Furthermore, Gourinchas noted that previous episodes have suggested that this trouble is front-loaded. In the early 1980s, following the Volcker disinflation, there were almost instant effects in EMDEs, with sharp increases in interest rates and, most notably, the Latin American debt crisis. In the 2013 taper tantrum, merely the announcement of tightening financial conditions elicited trouble in EMDEs.

Gourinchas challenged the authors to consider why, in the current circumstances of dollar appreciation and tightening financial conditions and given the precedent of front-loaded trouble in emerging markets, we have not yet seen major crises in emerging markets. Gourinchas wondered if this turmoil could be yet to come or if there had been improvements in how EMDEs have been dealing with the circumstances of dollar appreciation, for example, the early tightening of monetary policies, ahead of AEs, in Latin America.

Arvind Krishnamurthy, in response to Gourinchas's query, noted that India is currently going through foreign exchange reserves at a fast rate, suggesting that there may be signs of beginning trouble in EMDEs.

Gian Maria Milesi-Ferretti added to this discussion by positing three potential factors driving this episode of dollar appreciation. First, there are domestic factors of the United States, excess demand, inflation pressures, and interest rates. Second, there are massive negative shocks coming from Europe. As the paper considers the exchange rate of the dollar relative to other AEs, some of the observed dollar appreciation could be a product of war in Europe and trade shocks from segmented energy markets, which, Milesi-Ferretti explained, would be less indicative of real trouble for EMDEs. The third driving factor Milesi-Ferretti discussed was increasing global risk aversion.

Maurice Obstfeld responded, providing context for multiple important factors. First, Obstfeld observed that in the current episode EMDEs have been more proactive in raising interest rates in comparison to AEs, pointing to Brazil as an example of early and dramatic increases in interest rates. Second, Obstfeld noted that in the current situation the dollar had appreciated approximately 12 percent since mid-2021, through August 2022, which, while significant, is not as great as previous episodes. For example, from mid-2014 to the end of that appreciation episode, over an equivalent length

1. Julian T. Chow, Florence Jaumotte, Seok G. Park, and Yuanyan S. Zhang, *Spillovers from Dollar Appreciation* (Washington: International Monetary Fund, 2015); Şebnem Kalemli-Özcan, "U.S. Monetary Policy and International Risk Spillovers," working paper 26297 (Cambridge, Mass.: National Bureau of Economic Research, 2019), https://doi.org/10.3386/w26297; M. Ayhan Kose, Csilla Lakatos, Franziska Ohnsorge, and Marc Stocker, "The Global Role of the U.S. Economy: Linkages, Policies and Spillovers," working paper, Social Science Research Network, February 10, 2017, https://doi.org/10.2139/ssrn.2914672.

of time, the appreciation of the dollar was much greater, approximately 20 percent, breeding turmoil in China and other EMDEs. Considering the current appreciation as modest in comparison may clarify why substantial turmoil has not yet been seen in EMDEs. Lastly, Obstfeld suggested that the United States likely has further to go with interest rate increases, as the labor market has not yet slowed, high-yield spreads have not risen much, and the housing market has only slowed slightly. As these factors develop with further attempts from the Federal Reserve to reduce inflation, Obstfeld posited that more turmoil may be seen in EMDEs. Haonan Zhou concluded this discussion, mentioning a 1985 paper by Guillermo Calvo which argues that it is not the absolute level of the dollar exchange rate but the rate of appreciation that matters.[2]

Another thread of discussion considered the responsibilities of the United States given the wide impacts of US monetary policy. Jason Furman considered that foreign central bankers could approach the Federal Reserve and utilize this paper's research to advocate on account of the effects the dollar and US monetary policy have on their economies. Following, Furman asked if there were any elements of this paper that should encourage the Federal Reserve to care more about considering effects on EMDEs than they have before.

Building on this, Hanno Lustig remarked on the United States' exorbitant privilege, as the United States acts as the world's safe asset supplier and earns large convenience yields on Treasuries. Relatedly, Lustig continued, foreign issuers are incentivized to borrow in US dollars to capture some of this convenience yield. Lustig considered that as the dollar appreciates a problem is created for foreign borrowers, as the cost of servicing their debt increases. Trends of expansionary monetary policy and abrupt contraction especially create a problem. In conclusion, Lustig questioned if the exorbitant privilege of the United States in this position demands responsibility from US monetary policymakers to consider effects on the rest of the world. Additionally, Lustig noted the difficulty of the US-centric nature of the field, considering that the United States is an ineffective benchmark in macroeconomics as most countries are far more limited in what they can do in terms of fiscal policy.

Obstfeld responded briefly, commenting that in the past the Federal Reserve has considered effects on EMDEs. For example, after the lift-off

2. Guillermo A. Calvo, "Reserves and the Managed Float: A Search for the Essentials," *Journal of International Money and Finance* 4, no. 1 (1985): 43–60, https://doi.org/10.1016/0261-5606(85)90005-1.

rate hike in December 2015 the Federal Reserve waited to raise interest rates again for a full year. Obstfeld argued that a substantial motivation for this delay was the potential effects on turmoil in China and other EMDEs.

Reflecting on the importance of the dollar, Krishnamurthy raised a question on the safety of the Treasury. Analogizing Treasuries and safe dollar assets to gold stores, Krishnamurthy questioned what would occur if Treasuries were no longer safe or liquid and substantially lost value, comparing this to a world where half the gold suddenly disappeared. In response, Obstfeld considered this an interesting query on the confidence in the dollar's global role but noted that this subject would necessitate a lengthy discussion.

Additionally, Krishnamurthy pointed out that this paper added to evidence of strong spillover effects from the global financial cycle and global dollar cycle. This, Krishnamurthy remarked, challenged traditional international macroeconomic models, which contend that there should be no spillover effects with floating exchange rates.

In response to the discussants' comments from Şebnem Kalemli-Özcan and Matteo Maggiori, Obstfeld remarked on the need to consider the endogeneity of policy regimes and policies. Obstfeld expressed that a barrier to being able to know if interventions and capital controls are effective is that these policy actions are triggered by events, making controlled experiments difficult or impossible. Additionally, in response to Maggiori's discussion of causally identifying the effect of the dollar on the real economy, Zhou noted that some existing research has investigated this, such as research showing that dollar appreciation could causally increase the borrowing cost of syndicated loans.[3] Zhou added that this would be an important area for further research.

Lastly, Zhengyang Jiang asked about the potential need to include a default risk term in the model of currency excess return. Obstfeld clarified that in the paper the uncovered interest rate parity (UIP) condition referred to the dollar versus other AEs. In Kalemli-Özcan's comment, she discussed the potential for a UIP condition focused on the dollar versus EMDEs. Obstfeld explained that the inclusion of a default risk term might be appropriate in Kalemli-Özcan's UIP condition, but not in that of this paper.

3. Ralf R. Meisenzahl, Friederike Niepmann, and Tim Schmidt-Eisenlohr, "The Dollar and Corporate Borrowing Costs," International Finance Discussion Paper 1312 (Washington: Board of Governors of the Federal Reserve System, 2021), https://doi.org/10.17016/IFDP.2021.1312.