PUBLISHED FOR THE BROOKINGS INSTITUTION BY
JOHNS HOPKINS UNIVERSITY PRESS

Brookings Papers ON ECONOMIC ACTIVITY

BPEA
SPRING
2022

BROOKINGS

Janice Eberly and James H. Stock, Editors

Brookings Papers
ON ECONOMIC ACTIVITY

SPRING 2022

Brookings Papers

ON ECONOMIC ACTIVITY

SPRING 2022

JANICE EBERLY
JAMES H. STOCK
Editors

PUBLISHED FOR THE BROOKINGS INSTITUTION BY
JOHNS HOPKINS
UNIVERSITY PRESS

ISSN 0007-2303
E-ISSN 1533-4465

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

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

Brookings Papers
ON ECONOMIC ACTIVITY

SPRING 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. Working drafts of the papers are presented and discussed at conferences held twice each year, and the final versions of the papers and comments along with summaries of the general discussions are published in the journal several months later. Research findings are described in a clear and accessible style to maximize their impact on economic understanding and economic policymaking; the intended audience includes analysts from universities, governments, and businesses. Topics covered by the journal include fiscal and monetary policy, consumption and saving behavior, business investment, housing, asset pricing, labor markets, wage and price setting, business cycles, long-run economic growth, the distribution of income and wealth, international capital flows and exchange rates, international trade and development, and the macroeconomic implications of health care costs, energy supply and demand, environmental issues, and the education system.

We would like to thank the supporters of the *BPEA* conference and journal, including the Alfred P. Sloan Foundation; General Motors Company; the National Science Foundation, under grant no. 2048708; 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.

The views expressed by the authors, discussants, and conference participants in *BPEA* are strictly those of the authors, discussants, and conference participants and not of the Brookings Institution. As an independent think tank, the Brookings Institution does not take institutional positions on any issue. Conference drafts and recordings of the Spring 2022 *BPEA* conference can be accessed at https://www.brookings.edu/events/bpea-spring-2022-conference/.

CALL FOR PAPERS

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

ACCESSING THE JOURNAL

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

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

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

Henry Aaron *Brookings Institution*
Stephanie Aaronson *Brookings Institution*
Joseph Abadi *Federal Reserve Bank of Philadelphia*
Katharine Abraham *University of Maryland*
George Akerlof *Georgetown University*
Stefania Albanesi *University of Pittsburgh*
Alan Auerbach *University of California, Berkeley*
Laurence Ball *Johns Hopkins University*
Alina K. Bartscher *Danmarks Nationalbank*
Susanto Basu *Boston College*
Joe Beaulieu *Brevan Howard*
Ben Bernanke *Brookings Institution*
Sandra Black *Columbia University*
Olivier Blanchard *Peterson Institute for International Economics*
Alan Blinder *Princeton University*
Barry Bosworth *Brookings Institution*
William Brainard *Yale University*
Steven Braun *Council of Economic Advisers*
Jesse Bricker *Federal Reserve Board of Governors*
Ralph Bryant *Brookings Institution*
Elaine Buckberg *General Motors*
Jeremy Bulow *Stanford Graduate School of Business*
Gary Burtless *Brookings Institution*
David Byrne *Federal Reserve Board of Governors*
Elizabeth Cascio *Dartmouth College*
Gabriel Chodorow-Reich *Harvard University*
Gerald Cohen *Kenan Institute of Private Enterprise*
William A. Darity Jr. *Duke University*
Steven Davis *University of Chicago Booth School of Business*
Ryan Decker *Federal Reserve Board of Governors*
Brad Delong *University of California, Berkeley*
Jane Dokko *Federal Reserve Bank of Chicago*
Kathryn Dominguez *University of Michigan*
John Driscoll *Federal Reserve Board of Governors*
Karen Dynan *Harvard University*
Janice Eberly *Northwestern University*
George Eckerd *JPMorgan Chase*
Wendy Edelberg *Brookings Institution*
Kathryn J. Edin *Princeton University*
Barry Eichengreen *University of California, Berkeley*
Bruce Fallick *Federal Reserve Bank of Cleveland*
John Fitzgerald *Bowdoin College*
Benjamin Friedman *Harvard University*
Jason Furman *Harvard University*
William Gale *Brookings Institution*

Karen Glenn *Social Security Administration*
Claudia Goldin *Harvard University*
Austan Goolsbee *University of Chicago Booth School of Business*
Olga Gorbachev *University of Delaware*
Robert Gordon *Northwestern University*
Tracy Gordon *Urban Institute*
Gary Gorton *Yale School of Management*
Stephen Goss *Social Security Administration*
Josh Gotbaum *Brookings Institution*
Carol Graham *Brookings Institution*
Adam M. Guren *Boston University*
Joseph Gyourko *University of Pennsylvania Wharton School*
Robert Hall *Stanford University*
John Haltiwanger *University of Maryland*
Zhiguo He *University of Chicago*
Peter Henry *New York University Stern School of Business*
Brad Hershbein *W. E. Upjohn Institute for Employment Research*
Harry Holzer *Georgetown University*
Jessica Howell *College Board*
Caroline Hoxby *Stanford University*
Jennifer Hunt *Rutgers University*
Yannis Ioannides *Tufts University*
Lawrence Katz *Harvard University*
Jeffrey Kling *Congressional Budget Office*
Donald Kohn *Brookings Institution*
Arvind Krishnamurthy *Stanford University*
Moritz Kuhn *University of Bonn*
Robert Litan *Brookings Institution*
Deborah Lucas *Massachusetts Institute of Technology*
Maria Luengo-Prado *Federal Reserve Bank of Boston*
Byron Lutz *Federal Reserve Board of Governors*
Igor Makarov *London School of Economics*
N. Gregory Mankiw *Harvard University*
Robert McClelland *Urban-Brookings Tax Policy Center*
Bruce Meyer *University of Chicago*
Gian Maria Milesi-Ferretti *Brookings Institution*
Jeffrey Miron *Harvard University*
Frederic Mishkin *Columbia University*
Robert Moffitt *Johns Hopkins University*
Benjamin Moll *London School of Economics*
Adele Morris *Federal Reserve Board of Governors*
Vikram Nehru *Johns Hopkins University*
Maurice Obstfeld *University of California, Berkeley*
Jane Olmstead-Rumsey *Federal Reserve Bank of Minneapolis*
Karen Pence *Federal Reserve Board of Governors*
George Perry *Brookings Institution*
Thomas Philippon *New York University Stern School of Business*
Jonathan Pingle *UBS*
Richard Portes *London Business School*
Eswar Prasad *Cornell University and Brookings Institution*
Valerie Ramey *University of California, San Diego*
Alessandro Rebucci *Johns Hopkins University*
Tristan Reed *World Bank Group*
Carmen Reinhart *World Bank Group*

ALINA K. BARTSCHER
Danmarks Nationalbank

MORITZ SCHULARICK
University of Bonn

MORITZ KUHN
University of Bonn

PAUL WACHTEL
New York University Stern School of Business

Monetary Policy and Racial Inequality

ABSTRACT This paper aims at an improved understanding of the relationship between monetary policy and racial inequality. We investigate the distributional effects of monetary policy in a unified framework, linking monetary policy shocks both to earnings and wealth differentials between Black and white households. Specifically, we show that, although an accommodative monetary policy shock increases employment for Black households more than for white households, the overall effects on earnings are small. At the same time, an accommodative monetary policy shock has large effects on the wealth difference between Black and white households, because Black households own fewer assets that appreciate in value. This suggests an important trade-off if policymakers aim to reduce racial inequalities.

> The Fed has a profound impact on our economy. . . . Its existing mandate promotes maximum employment, and stable prices. . . . The Fed should add to that responsibility, and aggressively target persistent racial gaps in jobs, wages, and wealth.
>
> —Joseph Biden, "Racial Equity Plan Speech," July 28, 2020

> With regard to the employment side of our mandate, our revised statement emphasizes that maximum employment is a broad-based and inclusive goal. This change reflects our appreciation for the benefits of a strong labor market, particularly for many in low- and moderate-income communities.
>
> —Jerome H. Powell, "New Economic Challenges and the Fed's Monetary Policy Review," August 27, 2020

Conflict of Interest Disclosure: This research was supported by the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy—EXC 2126/1—39083886. Alina K. Bartscher is an employee of Danmarks Nationalbank. 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 currently not an officer, director, or board member of any organization with an interest in this paper.

Brookings Papers on Economic Activity, Spring 2022: 1–47

The racial tensions that spread across the United States in 2020 attracted the attention of monetary policymakers. Fifty years past the accomplishments of the civil rights movement, racial gaps in income and wealth remain enormous. There is widespread recognition that despite a decline in overt labor market discrimination and gains in educational opportunities since the onset of the civil rights movement, racial gaps persist and have even grown larger by some measures (Bayer and Charles 2018; Dettling and others 2017; Kuhn, Schularick, and Steins 2020; Thompson and Suarez 2017; Wolff 2017). The size and persistence of the gaps between both the income and wealth of Black and white households are striking (Chetty and others 2020; Emmons 2020). According to the 2019 Survey of Consumer Finances (SCF), the median wealth of a white household was $181,400, compared to only $20,700 for the median Black household, implying that the typical Black household owns only about 11 percent of the wealth of the typical white household. The income ratio is smaller but still large: the median income for Black households ($38,700) is 58 percent of the median income for white households ($67,200).[1]

Traditionally, macroeconomists and monetary policymakers held the view that racial inequities were outside their purview. However, the view that central banks should pay attention to racial inequalities in income and wealth has recently gained ground. For instance, Raphael Bostic (2020), president of the Federal Reserve Bank of Atlanta, suggests that the Federal Reserve "can play an important role in helping to reduce racial inequities and bring about a more inclusive economy" (par. 8). Yet, so far, we lack a deeper understanding of the impact monetary policy has on racial inequities, a topic that has for a long time received little attention from the research community. Our goal in this paper is to examine the effects of monetary policy on the income and wealth of Black and white households.

One line of thinking that links monetary policy to distributional outcomes runs as follows: at the business cycle frequency, a more accommodative monetary policy lowers unemployment and increases labor income for workers who would otherwise have become or stayed unemployed. Marginal workers who are drawn into the labor market by such policies are often from low-income and minority households. Consequently, the gap

1. Board of Governors of the Federal Reserve System, "Survey of Consumer Finances (SCF)," https://www.federalreserve.gov/econres/scfindex.htm.

between unemployment rates of Black and white households can be expected to shrink under a more accommodative policy.[2] In support of this view, Carpenter and Rodgers (2004) find a higher sensitivity of Black workers' labor market outcomes to monetary policy shocks. Coibion and others (2017) call this effect on low-income workers the *earnings channel*.

At the same time, monetary policy affects heterogeneous household balance sheets through its impact on asset prices (Brunnermeier and Sannikov 2012; Kaplan, Moll, and Violante 2018). Asset price changes will affect the racial wealth distribution if portfolios differ systematically between Black and white households. Using SCF data, we show that portfolio heterogeneity is a very pronounced fact in the data: Black households hold substantially different portfolios and in particular fewer financial assets than white households, so that monetary policy shocks potentially have larger effects on white households' portfolios. The median Black household has no stock holdings nor owns a house. Thus, any effect that monetary policy has on the price of such assets bypasses the majority of Black households. The effects could be particularly pronounced in the case of unconventional monetary policy, which explicitly aims at affecting asset prices (Bernanke 2020; Wu and Xia 2016).

In addition to the earnings and portfolio effects, monetary policy will have an impact on interest rates and dividends directly. We call the effect on interest earnings on savings and bonds, dividend earnings and the gains or losses from mortgage refinancing the *capital income effect*. To the extent that Black and white households' portfolios differ, there will be differential capital income effects of monetary policy.

Since accommodative monetary policy boosts asset returns, it is likely that the portfolio and earnings effects go in opposite directions. On the one hand, more accommodative monetary policy may benefit Black households by reducing unemployment and increasing labor market participation and earnings, thereby helping to reduce the racial income gap—and, over time, even the wealth discrepancy if part of the additional income is saved.

2. This channel is often emphasized in policy discussions (Aliprantis and Carroll 2019). In the words of Atlanta Fed president Raphael Bostic (2020): "The Federal Reserve acts to create a foundation upon which businesses, families, and communities can thrive. Our success means that businesses can grow faster and hire more workers and that more innovation can be supported, which would mean more opportunities for African Americans and others who have not been as attached to the economy" (par. 9).

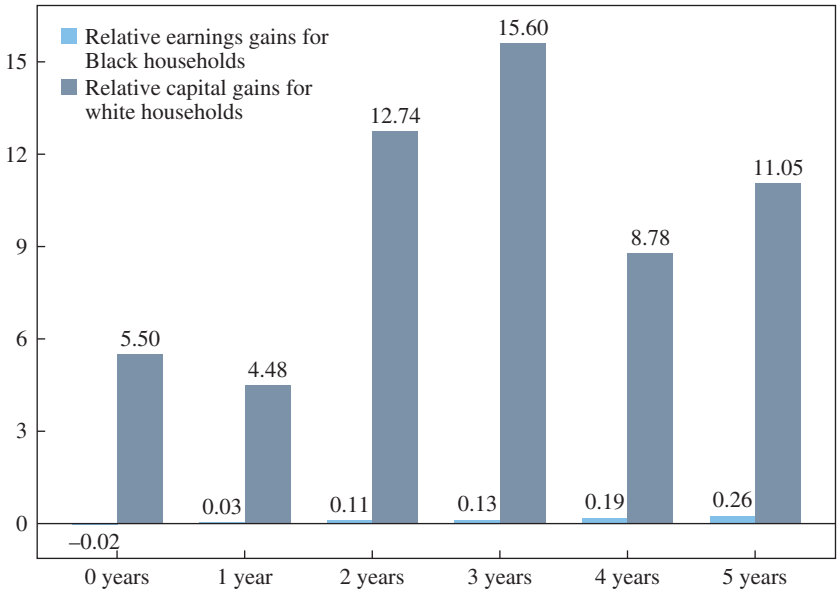
On the other hand, the same policies may widen racial wealth differences if white households benefit more from rising asset prices than Black households due to their different portfolio composition and greater wealth. The capital income effects can go in either direction, since lower interest rates reduce household interest income but the opportunity to refinance mortgages at a lower rate can have positive effects on disposable income.

This paper quantifies and compares the size of the earnings, portfolio, and capital income effects of monetary policy. We begin with a comparative statics exercise, examining the impact of a given change in asset prices and interest rates. We then develop a unified empirical framework that uses instrumental variable local projections following Stock and Watson (2018) and Jordà, Schularick, and Taylor (2020) to study the effects of a monetary policy shock on asset prices, interest rates, and Black-white employment gaps over a five-year horizon. For this analysis, we rely on the most widely used monetary policy shock series—the (extended) Romer-Romer shocks (Coibion and others 2017). We apply the asset price and interest rate changes to the portfolios of white and Black households from the most recent SCF wave in 2019 and determine the effect on the net wealth of Black and white households. We further combine the estimated effects on the unemployment gap with unemployment and earnings data from the SCF and compare them to the portfolio effects in response to the estimated monetary policy shocks over different time horizons.

KEY FINDINGS A 100 basis point accommodative monetary policy shock leads to larger employment gains for Black households but also to larger wealth gains for white households. More precisely, the Black unemployment rate falls more than the white unemployment rate after an unexpected accommodative interest rate shock. This translates into a relative earnings gain for the mean Black household relative to the mean white household. Our results indicate that after five years the cumulative earnings gain for Black households relative to white households is $134.

The same monetary policy shock pushes stock and house prices up, while lowering bond yields and increasing dividend payments. Since the average wealth of white households is about six times that of Black households, and moreover the composition of Black and white portfolios is markedly different, there are large differences in the effects on the wealth of Black and white households. For white households, we find that on average, a 100 basis point accommodative policy shock leads to capital gains from asset price changes of $18,900 after five years, which is almost one-fifth of their average annual income. The wealth gains for Black households

Figure 1. Comparison of Relative Earnings and Portfolio Effects



Source: Authors' calculations.

Note: The graph compares the cumulative relative earnings effect for Black households to the relative portfolio effect for white households based on an expansionary 100 basis point monetary policy shock. The effects are reported as a percentage share of average annual household income of the respective racial group. See section V.D for the calculation of the relative earnings effect. The relative portfolio effect is the difference between the capital gains of white and Black households from figure 10.

are substantially smaller, about $3,300, corresponding to 6 percent of their average annual income.

In figure 1, we show the year-by-year accumulated earnings and portfolio effects as a percentage of each group's income. For an easier comparison, we constructed the differences to be always positive. Capital gains are larger for white households and earnings gains are larger for Black households. The earnings effect is the relative earnings gains for Black households, and the portfolio effect is the relative capital gains for white households.[3] The details on how these effects are calculated can be found in section V.E.

3. The earnings effect applies to a flow, while the portfolio effects reflect capital gains on the stock of wealth. To take this into account, we accumulated the differential earnings effects over time.

Even as the earnings effect accumulates over time, it remains orders of magnitude smaller than the effects from capital gains.

While the earnings effect increases the consumption possibilities of households directly, capital gains need to be realized first. To make a fair comparison of the wealth and earnings effects, we thus look at the consumption effects of capital gains. Typical estimates indicate that the marginal propensity to consume (MPC) out of capital gains is about 3 percent. This means that our estimated difference between the capital gains received by white and Black households after five years, about $15,600, leads to additional consumption expenditures of $470 for white compared to Black households. The relative consumption effect of capital gains for white households in year 5 is three and a half times larger than the relative earnings gain for Black households cumulated over five years. An accommodative monetary policy would need to have a much larger effect on Black employment and earnings in order to match the impact of changes in asset prices on the consumption of white households.

Our empirical findings strongly suggest that monetary policymakers face a trade-off: monetary accommodation widens racial wealth inequality as it reduces income inequality. There is little reason to think that monetary policy can play a significant role in reducing racial inequities in both income and wealth at the same time. The conventional tools of monetary policy seem ill suited for these important tasks.[4]

STRUCTURE OF PAPER In section I, we briefly discuss prior literature on the channels of monetary policy and its distributional effects. In section II, we discuss racial inequalities in income and wealth, present the data, and discuss portfolio differences between Black and white households. In section III, we examine the effect of a 10 percent change in asset prices and a 100 basis point change in interest rates on the portfolios of Black and white households.

We present our estimates of the effects of a monetary policy shock on asset prices, interest rates, dividends, and the wage and unemployment gaps in section IV. In section V, we examine the impact of a typical monetary policy shock on Black and white wealth and capital income and compare the wealth effects to the estimated earnings effects. The last section concludes.

4. Our analysis is based on a surprise change to the federal funds rate; we do not examine the effects of nonconventional policies at the zero lower bound, such as quantitative easing.

I. Policy Channels and Empirical Literature

We begin with a brief discussion of the theoretical literature that relates monetary policy–induced asset price change to consumption. We then summarize the empirical literature on the distributional effects of monetary policy.

I.A. Policy Channels, Consumption, and Welfare

There are at least two notable channels through which monetary policy–induced asset price changes can affect consumption: wealth effects and collateral constraints.[5] First, changes in both house and stock prices can affect household consumption. For housing, Berger and others (2018) demonstrate that a calibrated heterogeneous agent model is quantitatively consistent with large estimated house price effects on consumption. Chodorow-Reich, Nenov, and Simsek (2021) study stock market wealth effects and also find significant effects on consumption.

In addition, capital gains can lead to redistribution and welfare effects when some households sell assets to realize their capital gains, as emphasized in Moll (2020). Households planning to buy assets that appreciate will experience welfare losses, while households who plan to sell will experience gains. For instance, households at different points of the life cycle differ in whether they plan to buy or sell assets (Greenwald and others 2021). Glover and others (2020) explore such life cycle redistribution with a focus on the consequences of the large asset price changes during the financial crisis. A similar logic can be applied to racial differences in asset holdings. If past discrimination in housing markets implies that Black households are structurally "short" in housing and have plans to become homeowners, asset price increases would tend to make those households worse off.

Second, rising asset prices may also temporarily relax collateral constraints and facilitate borrowing for housing or entrepreneurial investments. To the extent that such effects fall predominantly on white households, they can induce permanent effects on income and wealth inequality (Boerma and Karabarbounis 2021). Similarly, after an accommodative monetary

5. Recent theoretical macro models with heterogeneous agents have emphasized the asset price channel of monetary policy transmission (Auclert 2019; Auclert, Rognlie, and Straub 2020; Caramp and Silva 2021; Kekre and Lenel 2020).

policy shock households can permanently lock in lower mortgage rates through refinancing. The evidence we present below is consistent with such permanent gains accruing predominantly to white households.

This discussion implies that even if monetary policy shocks only have temporary effects on asset prices, they can have persistent economic consequences. Moreover, our estimated effects of a policy rate shock on asset prices remain visible over a multiyear period, as in other recent research (Paul 2020). Hence, even temporary policy shocks can alter the equilibrium characteristics of the economy with long-lasting effects.

I.B. Empirical Literature

There is a large body of literature on the distributional consequences of monetary policy that mostly focuses on income and consumption inequality.[6] Coibion and others (2017) find that a contractionary monetary policy shock increases inequality in pretax incomes and consumption. They estimate the effects of monetary policy shocks in the spirit of Romer and Romer (2004) on aggregate inequality measures. Using a similar approach and administrative data from Norway, Holm, Paul, and Tischbirek (2021) find evidence that contractionary monetary policy shocks increase inequality in disposable income and consumption but decrease wealth inequality. By contrast, Andersen and others (2021) find an increase in disposable income inequality after an accommodative monetary policy shock. They use Danish micro data and exploit the peg of the Danish krone to the euro in order to identify monetary policy shocks. Unlike the previous two studies, the authors consider household income distribution within age groups and estimate inequality effects based on income effects at the household level, instead of estimating the effects on aggregate measures of inequality. While Andersen and others (2021) find monotonically increasing effects of accommodative monetary policy shocks on disposable incomes along the income distribution, Amberg and others (forthcoming) find *U*-shaped income effects based on Swedish administrative data. They identify monetary policy shocks with a high-frequency approach and study the effects on total posttax income. Similar to Andersen and others (2021), they compute inequality effects from income effects at the individual level. Due to the *U*-shape of income effects, the overall effects on income inequality depend on the inequality measure considered. For instance, they find that inequality increases as measured by the

6. See Colciago, Samarina, and de Haan (2019) for an overview.

Gini coefficient, yet decreases as measured by the ratio of the 90th to the 10th percentile.

Only a few papers have explicitly focused on the effect of monetary policy on wealth inequality. Adam and Tzamourani (2016) use euro area data from the Household Finance and Consumption Survey to estimate the impact of changes in different asset prices along the wealth distribution. Albert and Gómez-Fernández (2021) use the high-frequency monetary policy shocks of Gertler and Karadi (2015) in a structural vector autoregressive (VAR) model with US data to estimate the effects on interest rates, dividends, and stock and house prices. They link these effects to data from the 2016 SCF and find that an expansionary monetary policy shock increases wealth inequality, especially in the long run. Lenza and Slacalek (2021) examine the effect of quantitative easing shocks in the euro area on wealth and income distributions. They find little effect on the wealth distribution and a noticeable effect on the income distribution due to increased employment in lower-income households, although the effects are not long lasting. Mäki-Fränti and others (2022) study the effects of both conventional monetary policy and quantitative easing shocks on income and wealth inequality in Finland, finding positive but small effects on both income and wealth inequality for both types of shocks.

Although we are not aware of any other examination of the effect of monetary policy on the racial wealth gap, the size and persistence of the wealth gap has been shown in previous work, most recently by Emmons, Kent, and Ricketts (2019), Kent and Ricketts (2021), and Aladangady and Forde (2021). The differential effect of monetary policy on Black and white unemployment rates was observed in the 1990s; see, for example, Zavodny and Zha (2000). Carpenter and Rodgers (2004) find a higher sensitivity of Black workers' labor market outcomes to monetary policy shocks. Finally, Rodgers (2008) explores differential effects of monetary policy on the duration of unemployment for Black and white workers. His evidence points toward a stronger effect on the unemployment duration of Black workers than for white workers after contractionary monetary policy shocks.

II. Racial Inequalities in Income and Wealth

In this section, we describe the Survey of Consumer Finances (SCF) data and present summary statistics. The data from the 2019 SCF indicate that the median wealth of white households was almost nine times higher than for Black households, while white median income was 1.7 times greater than for Black households. Not only is the wealth gap between Black and

white households large, it has hardly changed over the last fifty years. We show trends in the financial situation of Black and white households with the data compiled by Kuhn, Schularick, and Steins (2020) from early waves of SCF going back to 1950.

II.A. SCF Data

The SCF provides representative data on the financial situation of US households, employing a survey design that oversamples wealthy households. The detail of the financial information, the data quality, and the extent of the household coverage have made the SCF the primary source for studying the distribution of income and wealth among US households. In the 2019 SCF data, 68 percent of household heads reported being white, 16 percent reported being non-Black and non-white, and 16 percent of households reported that they have a Black head of household. For our analysis, we focus on households who either have a Black or a white head.[7]

We follow the definitions of income and wealth in previous literature (Bricker and others 2016; Kuhn and Ríos-Rull 2016; Kuhn, Schularick, and Steins 2020). In particular, wealth is the sum of all assets minus all debt of a household. We consider marketable wealth so that we do not include claims against Social Security or defined-benefit retirement plans. Defined-contribution retirement plans are part of marketable wealth and constitute 17 percent of wealth in the United States (Kuhn and Ríos-Rull 2016). Housing includes the primary residence, other residential real estate, and the net value of nonresidential real estate. For income, we consider income from all sources; for earnings, we use wage and salary income. We convert all nominal variables throughout the paper to 2019 dollars using the Consumer Price Index (CPI).

We use the approach of Bricker and others (2017) to construct household holdings of all asset classes, calculating total stock and bond positions as the sum of direct and indirect holdings. Directly held bond and stock investments are allocated to their respective positions. For indirect holdings, we allocate stock and bond investment components for stock and bond mutual funds, annuities and trusts, retirement accounts, and investment savings accounts to the respective total stock and bond holdings. In the end, total stock holdings are the sum of directly held stocks, stock mutual funds (where we take 50 percent of the holdings of combination mutual

7. The SCF convention is that in a couple the male spouse is the head of household, and we follow this convention.

Table 1. Mean and Median Black and White Wealth and Income in the 2019 SCF

	Means		*Medians*		*Share with holdings (%)*	
	White	*Black*	*White*	*Black*	*White*	*Black*
Bonds	122,700	19,600	0	0	47	27
Housing	353,500	104,700	170,000	0	75	46
Equity	474,000	40,900	9,000	0	64	35
Other nonfinancial assets	33,400	13,500	17,000	8,000	90	72
Liquid assets	57,000	13,900	8,000	1,400	99	95
Other financial assets	28,400	7,600	0	0	37	30
Net wealth	951,300	139,800	181,400	20,700		
Debt	117,300	60,400	35,000	10,100		
Income	113,300	58,100	67,200	38,700		

Source: Authors' calculations.

Note: All dollar values are rounded to the nearest $100. Housing includes other real estate. Equity includes business wealth. Nonfinancial assets are the value of vehicles and other nonfinancial assets (e.g., jewelry or gold). Liquid assets are the sum of checking accounts, saving accounts, call accounts, money market deposit accounts, prepaid accounts, and certificates of deposit. Other financial assets include the cash value of life insurance.

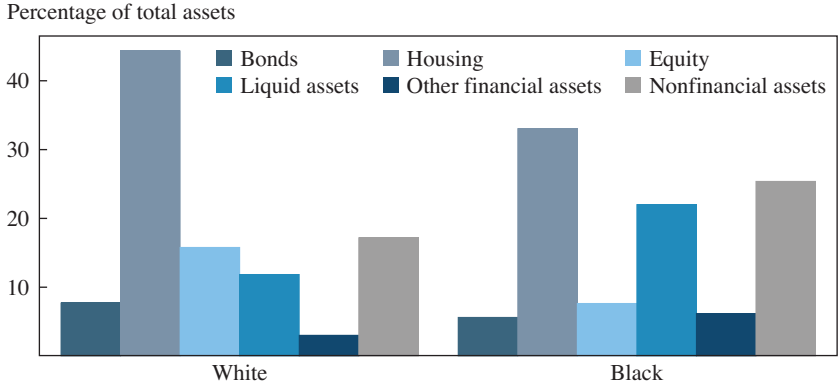
funds), and the share of retirement plans, other managed investments, and investment saving accounts which are invested in stocks, as reported by the survey participants. We proceed accordingly for bonds.

2019 SCF SUMMARY STATISTICS Table 1 provides a summary of the financial situation of Black and white households in the United States in 2019. We report several asset components from household balance sheets, as well as total debt, wealth, and income. We report means and medians for asset positions, wealth, debt and income, and in addition the share of households with positive holdings of each asset class.

The SCF data show that the average Black household has 51 cents for each dollar of white household income. The average wealth gap is dramatically larger; the average Black household has only 15 cents per dollar of white household wealth. The racial wealth gap is prevalent on the entire household balance sheet but it is much smaller for nonfinancial assets. For example, for housing, the average Black household owns 30 cents per dollar of the average white household. By contrast, if we look at equities, Black households hold on average only 9 cents for every dollar of equity held by white households.

Comparing means and medians highlights the large skewness of the US wealth distribution, with means being much larger than medians. The racial wealth gap is larger at the median than at the mean, with the typical Black household owning only about 11 percent of the wealth of the

Figure 2. Average Portfolio Shares of White and Black Households



Source: Authors' calculations.

typical white household.[8] For many asset types, the median is zero or close to zero because the share of households with holdings is small. The last two columns of table 1 show that only 35 percent of Black households own equities, just a bit more than half the share of white households. Black households are heavily underrepresented at the top of the US wealth distribution, where financial wealth is concentrated (Kuhn and Ríos-Rull 2016). Many Black households in the United States do not have any financial assets at all, so if asset prices increase, they will not benefit.

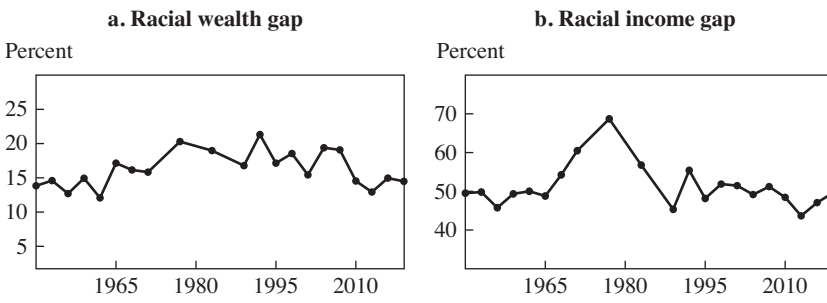
Figure 2 displays the portfolio composition of Black and white households by showing the average share of each asset class in total assets.[9] Housing is the largest portfolio component for both Black and white households. The housing share is larger for white households, who on average hold 44 percent of their assets in housing, compared to an average share of 33 percent for Black households.

The equity share of white households (around 16 percent) is about twice as high as for Black households. For bonds, the discrepancy in average portfolio shares between Black and white households is smaller. Differences in portfolio composition translate into differences in exposure to asset

8. Medians are computed within asset classes and therefore might not correspond to the asset holdings of the median-wealth household.

9. Note that the figure shows average portfolio shares, which differ from the portfolio shares of the average household obtained by dividing the average holdings of each asset class by average total assets (as found in table 1). The latter would amount to an asset-weighted average of the household-level portfolio shares.

Figure 3. Long-Run Trends of the Racial Wealth and Income Gaps



Source: Authors' calculations.

Note: Panels show the evolution of the ratio of average Black to average white wealth and income over time. The data were Winsorized at the 1st and 99th percentile within each year-race bin.

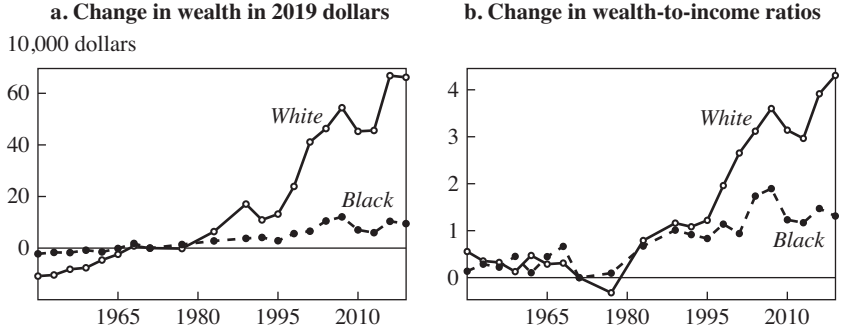
price changes (Kuhn, Schularick, and Steins 2020). The portfolio shares for housing, equities, and bonds are larger for white households, making them more exposed to changing asset prices than Black households, who have a larger share of low-return liquid assets, life insurance, and nonfinancial assets such as vehicles.

II.B. Trends in Racial Income and Wealth Inequality

We use data from Kuhn, Schularick, and Steins (2020) to show trends in racial wealth and income gaps since the 1950s. Based on these data, figure 3, panel a, shows the racial wealth gap, that is, the ratio of average Black to average white wealth, and panel b shows analogous results for income. The racial wealth gap decreased somewhat from the 1970s until the 2008 financial crisis; it now stands at about 15 percent, just as in the 1950s. This reversal was largely driven by the collapse of house prices (Kuhn, Schularick, and Steins 2020; Wolff 2016). In particular, despite some fluctuations over time, the ratio of Black to white average stock and business wealth has remained at persistent low levels, without any indication of an upward trend (see online appendix figure A.1, panel a). The same holds true for the second major asset class, namely, housing. The housing wealth gap only closed for a short period in the 2000s (see online appendix figure A.1, panel b). The trends in the income gap are similar. There was a reduction in racial income inequality since the mid-1960s, which was followed by a return to earlier levels of the gap in the 1990s.[10]

10. Similar patterns emerge when looking at medians, although the median gaps in income and wealth are slightly smaller in the 2010s than in the 1950s.

Figure 4. Change in Wealth and Wealth-to-Income Ratios Relative to 1971



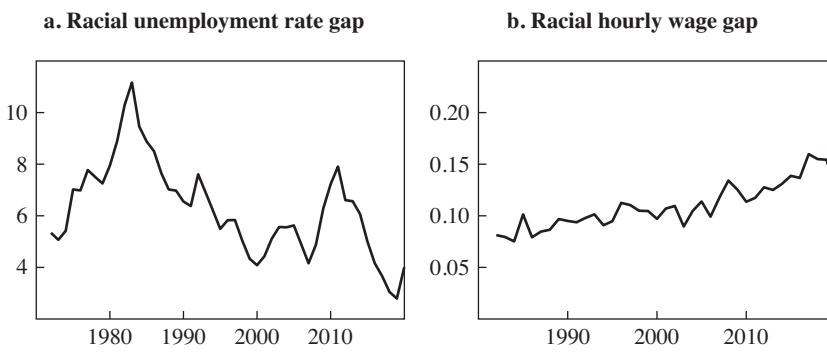
Source: Authors' calculations.
Note: Changes shown are the differences from the 1971 values.

In figure 4, panel a, we contrast the dollar changes in average wealth of Black and white households in the United States over the past seventy years relative to 1971. While average white wealth increased by about $650,000 in 2019 dollars, the wealth of Black households increased by a little more than $100,000, keeping the wealth gap at roughly the same level as in the 1950s. The stock market boom of the 1990s provided a boost to white wealth, which increased by about $400,000 per household between 1995 and 2007, while average Black wealth increased by less than $100,000. Such large differences stem from the much higher exposure to equity markets of wealthy, typically white, US households.[11]

Figure 4, panel b, compares the changes in wealth-to-income ratios of Black and white households relative to the 1971 ratio. We find a strong co-movement from the early 1950s to the mid-1990s, when a rapid divergence took place. By 2019, white households owned $8.4 of wealth per dollar of income, while Black households owned only $2.4 (see table 1). Between the early 1970s and today, Black households increased their wealth by only slightly more than one year's income, while the wealth of white households increased by about four times their annual income. This stark difference was mainly driven by equity and business wealth. Online appendix figure A.2 shows the counterfactual change in Black and white wealth-to-income ratios when keeping wealth from equity, businesses, and defined-contribution pension accounts (which are to a large extent

11. Increases in equity prices during the 1990s also tended to increase wealth inequality *among* white households (Kuhn, Schularick, and Steins 2020).

Figure 5. Racial Unemployment and Wage Gaps



Source: Authors' calculations based on FRED and Current Population Survey (CPS) outgoing rotation groups.

Note: The left panel shows the racial unemployment gap from 1972 to 2020 (in percentage points), computed as the difference between the average annual unemployment rates of Black and white workers. The right panel shows the racial (log) wage gap from 1982 to 2020 for annual averages of log wage data for Black and white workers who are paid by the hour.

invested in equities) fixed at their 1971 levels. Without the equity- and business-induced wealth gains, wealth-to-income ratios would have remained relatively stable from the mid-1990s to today, apart from a short-lived housing-based increase around the financial crisis. In particular, Black and white wealth-to-income ratios would have evolved in a strikingly similar way. High wealth-to-income ratios imply that changes in asset prices lead to large wealth gains relative to income. Accordingly, differences in saving rates, which operate on income flows, can have only a small impact on the wealth gap compared to the impact of asset price changes, which operate on much larger wealth stocks (Kuhn, Schularick, and Steins 2020).

UNEMPLOYMENT AND WAGE GAP TRENDS In addition to the large wealth and income differences between Black and white households, there are differences in labor market outcomes of Black and white households. Specifically, the racial gaps in unemployment rates and wages are large. The racial unemployment gap is the focus of discussions about the earnings effect of monetary policy. We use Bureau of Labor Statistics (BLS) data on unemployment rates starting in 1972, when Black unemployment rate data become available.[12] Figure 5, panel a, shows the Black-white annual unemployment gap from 1972 to 2020. The gap has rarely been smaller

12. The gap is the difference between Black and white unemployment rates, where the data are seasonally adjusted with Census X-12 ARIMA.

than 4 percentage points. It almost reached 12 percentage points during the 1982 recession and hit a low of 3 percentage points in the tight labor market prior to the COVID-19 pandemic.

For the wage gap, we use data for Black and white workers from the Current Population Survey (CPS) (Flood and others 2021). Wage data for all employed workers paid by the hour in the CPS outgoing rotation groups are available from January 1982 onward. The racial wage gap shown in figure 5, panel b, is the difference between log wages of Black and white workers. The series has an upward trend from 8 percent lower wages for Black workers in 1982 to almost 15 percent today. The wage gap does not show systematic cyclical fluctuations around this secular trend. The increasing wage gap counteracts some of the effects of the historical decline in the racial unemployment gap shown in figure 5, panel a.[13]

III. Household Portfolios, Asset Prices, and Interest Rates

In the following, we illustrate the different sensitivity of Black and white household asset portfolios to changes in asset prices and interest rates.

III.A. Portfolio Composition and Asset Price Changes

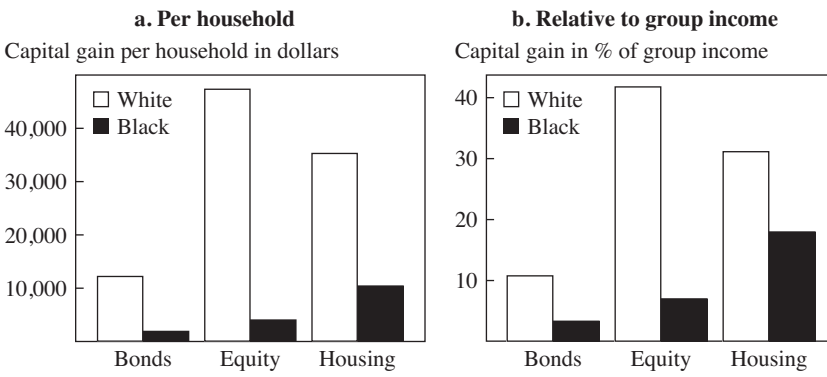
To illustrate the effect of asset price changes, we consider a 10 percent increase in the price of each asset and look at how this affects the wealth of the average Black and white household.[14] Figure 6, panel a, shows the dollar wealth changes for three major asset classes—bonds, equity, and housing—following a 10 percent asset price increase. Changes in asset prices lead to much larger capital gains for white compared to Black households, which is not surprising given the large differences in the average wealth levels shown in table 1.

These racial differences in capital gains are only partially mitigated when we look at the wealth gains relative to household income, as shown in figure 6, panel b. Even in relation to income, we find the differences still to

13. Another reason for the trend might be changes in the group of workers who are paid by the hour. We also considered data on the racial gap in mean and median weekly earnings and found our results to be robust.

14. A 10 percent change seems to be a reasonable benchmark in light of the substantial increases in asset prices that have occurred during the past fifteen years. Over this time period, US home prices rose by 69 percent, stock prices by 95 percent, and bond prices by 22 percent. These numbers are based on the annual average S&P/Case Shiller US National Home Price Index, the end-of-year S&P 500 stock price index, and the annual average US ten-year government bond yield with the assumption that duration is seven years.

Figure 6. Capital Gains from 10 Percent Increase in Asset Prices



Source: Authors' calculations.

Note: Capital gains are computed as the product of the price change and the average stock of asset holdings of the respective racial group.

be large. For example, if stock prices rise by 10 percent, capital gains for white households are over 40 percent of annual income. For Black households, the corresponding number is less than 10 percent. These results mean that any capital gains from asset price changes accrue disproportionately to white households.

Housing, the largest asset of most Americans, is particularly important due to the possibility of racial discrimination in housing markets.[15] Table 1 already showed that homeownership rates are lower for Black households. Zero housing wealth at the median implies that not even every second Black household owns a house. By contrast, the housing wealth of white households at the median is already $170,000, more than the average total wealth of Black households. In online appendix III, we conduct counterfactual analyses where we equalize the homeownership rates of Black and white households or the house values of Black and white homeowners. We find that the average gap in capital gains relative to income between Black and white households would be almost closed if Black households had the same propensity to be homeowners as white households, and it would be more than closed if Black homeowners' houses were as valuable as those of white homeowners.

Whether it is redlining, other forms of discrimination, or other factors that have led to Black households owning fewer and less valuable homes,

15. See, for example, Zonta (2019) and Joint Center for Housing Studies (2020).

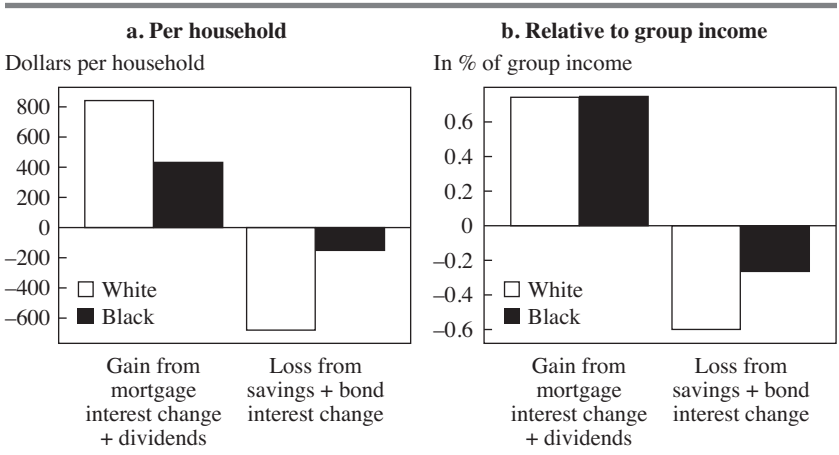
these differences mean that Black households gain less from overall home price appreciation. This potentially fuels further racial inequalities when monetary policy leads to capital gains in the housing market.

III.B. Portfolio Composition and Interest Rate Changes

Black and white households are also affected differently when interest rates and dividend payments change. Households are affected by such changes in several ways after an accommodative monetary policy shock. First, lower interest rates will lead to lower interest income on bank accounts and deposit-type assets. Unlike for fixed-rate bonds that will increase in value, the money value of an account balance will not change. What will change are the future income flows from this balance, making a household with a positive balance poorer in expectation. Falling interest rates also reduce the interest earnings on bonds when maturing bonds are reinvested at a lower rate. Around 13.4 percent of corporate and 20.6 percent of mortgage-backed bonds are refinanced each year, which leads to a loss in interest income when rates fall.[16] Second, we assume that a policy accommodation that leads to increased equity prices and profits will also lead to an increase in dividend payments. Given the higher stock market participation and average stock holdings of white compared to Black households, this source of income mainly matters for white households. The final way in which households are affected by lower interest rates is via borrowing, in particular if the household borrows with a mortgage contract that allows refinancing at a lower interest rate. Most US mortgages are fixed-rate mortgages with a built-in call option that allows for the opportunity to prepay. Although refinancing is costly and cumbersome, refinancing activity typically increases when interest rates fall. The lower rates will persist for the remaining duration of the mortgage (Bhutta and Keys 2016). Refinancing activity is therefore an important example where even transitory changes in interest rates resulting from monetary policy can have long-lasting redistributive effects as households lock in the lower interest rate for the duration of the mortgage. If the mortgage balance is not increased upon refinancing, but future interest payments are lowered, the household is effectively richer. In this sense, households with reduced monthly payments will be richer even if their net worth is unchanged in an accounting sense.

16. The proportions of bonds maturing are estimated as total issuance less the change in bonds outstanding as a fraction of bonds outstanding, averaged over the ten years since 2011, based on data from the Securities Industry and Financial Markets Association; https://www.sifma.org/resources/archive/research/statistics/.

Figure 7. Capital Income Effects from a Decline in Interest Rates after One Year



Source: Authors' calculations.

Note: The left panel shows the average gains for Black and white households after a 100 basis point decline in mortgage interest rates and a 1 percent increase in dividend income, and their average losses after a 100 basis point decline in savings and bond interest rates. The right panel shows the same gains and losses as a percentage share of each group's total income.

Exploring the capital income effects of a monetary accommodation through interest rate, dividend, and refinancing effects, given changing balances and maturities, is very complex. To examine the impact of monetary policy–induced interest changes, we will consider a 100 basis point fall in interest rates over a one-year horizon. First, we compute the loss in income from lower rates on deposit-type assets and refinanced corporate and mortgage-backed bonds for the one-year horizon. This effect is the foregone income due to the fall in interest rates. Second, to compute the effect from reduced mortgage payments, we assume that all fixed-rate mortgages are refinanced to the lower rate without changing the mortgage balance or remaining time to maturity. The latter effect reflects the change in annual mortgage payments if a household locks in the new lower interest rate by refinancing a fixed-rate mortgage. Finally, we consider a 1 percent increase in dividend incomes.

Figure 7, panel a, shows the average loss in interest income on liquid assets and newly issued bonds after a 100 basis point decline in interest rates and the average gain from mortgage refinancing and higher dividend incomes. Given that the average holdings of liquid assets and bonds are larger for white households (as shown in table 1), it is expected that the decline in interest income is much larger for white than for Black

households. Over one year, the interest income of the average Black household goes down by about $160, and it goes down about four times as much for white households. Expressing these losses relative to income, figure 7, panel b, shows that they are small: about 0.6 percent of annual income for white households and about half as much for Black households.

Mortgage debt balances of US households, after four decades of growth, are large and correspond to almost 100 percent of SCF household income (Bartscher and others 2021). The dollar decline in mortgage payments from refinancing after a 100 basis point decrease in interest rates is shown in figure 7, panel a, along with the gain from higher dividend incomes—which is, however, small. We find that mortgage payments per household decline by $800 for white households and by roughly half as much, $400, for Black households. Figure 7, panel b, shows that, as a fraction of current annual income, the responses are almost equal. For both Black and white households, the reduction in mortgage payments corresponds to roughly 0.7 percent of annual income. It is however important to keep the distribution of homeownership in mind; more than every second Black household does not own a house and therefore typically also does not owe mortgage debt. Moreover, the calculations are based on a scenario in which all households actually take advantage of the fall in the mortgage interest rate and refinance. Yet recent evidence by Gerardi, Willen, and Zhang (2021) suggests that Black households benefit less because they are substantially less likely to refinance when interest rates decline.

IV. Monetary Policy, Asset Prices, and the Unemployment Gap

In section II, we showed the heterogeneity in portfolio composition between Black and white households and differences in the racial unemployment and wage gaps. In section III, we showed that portfolio heterogeneity leads to different gains when an expansive monetary policy results in an increase in asset prices and dividends and a decline in interest rates. In the following, we will develop estimates of the effects of a monetary policy shock on the prices of assets—equities, houses, and bonds—as well as on interest rates, dividends, and labor market outcomes. In section V, we will combine these estimates with the household portfolio data from the SCF in order to investigate the wealth and capital income effects of an accommodative policy shock for Black and white households and compare them to the earnings effects that result from changes in the racial unemployment gap.

To study the effects of monetary policy shocks on asset prices and other outcomes, we use instrumental variable local projections following Stock

and Watson (2018) and Jordà, Schularick, and Taylor (2020). We employ the widely used extended Romer and Romer series for policy shocks (Coibion and others 2017; Romer and Romer 2004) as an instrument for the change in the federal funds rate. In the interest of comparability and transparency, we will also show simple local projection results for uninstrumented changes in the federal funds rate. Although there is a wide range of estimates of the effects of policy shocks on macroeconomic outcomes in the economic literature, we maintain that our estimates provide plausible approximations that illustrate the underlying economic mechanisms.

We show estimates of the impact of monetary policy shocks over a five-year period. There is a growing consensus in the literature that monetary policy moves asset prices over extended periods. Rigobon and Sack (2004) and Bernanke and Kuttner (2005) pioneered empirical approaches, finding substantial effects of policy surprises on stock prices that mainly come from changes in risk premia (excess returns). In both studies, a surprise 100 basis point shock lowers stock prices by between 5 and 7 percent. Jordà, Schularick, and Taylor (2015) document substantial effects of exogenous changes in monetary conditions on all major asset classes over multiyear horizons in a long-run cross-country data set. A recent paper by Paul (2020) argues that monetary policy today has larger and more persistent effects on asset prices than in the past. Similar findings have been reported for nonconventional monetary policy (Bernanke 2020; Wu and Xia 2016). The same mechanism that we describe in this paper—greater wealth effects for white households than for Black households following monetary policy–induced asset price gains—can be applied to these findings as well. Only the size and duration of the effects will vary across different studies.

IV.A. Estimation of the Effects of Monetary Policy

We treat the monetary policy shock measure as a proxy for the structural shocks in the instrumental variable local projections setup. The intuition is that surprises and structural shocks are imperfectly correlated. Monetary surprise measures suffer from measurement error due to noise and random zero observations in months without Federal Open Market Committee meetings. Instrumenting the federal funds rate instead of future rates also reduces the problems raised by the potential release of private central bank information (Nakamura and Steinsson 2018). Throughout the analysis, we scale the policy shocks to represent a 100 basis point surprise cut in the federal funds rate.

Let Δr_t denote the change in the federal funds rate at time t. We denote as x the vector of controls, which includes two lags of the outcome and the

interest rate variables, as well as other variables such as the unemployment rate, inflation, industrial production, corporate bond yields, the dividend-price ratio, money growth, and asset prices. Consider the following set of local projections relating future economic outcomes such as stock and house price changes, as well as the Black-white unemployment rate, to changes in interest rates today:

$$(1) \qquad y_{t+h} = \alpha_h + \Delta r_t \beta_h + x_t \gamma_h + \nu_{t+h}; \text{ for } h = 0, \ldots, H-1,$$

where $t = 1, \ldots, T$.

Estimates of this equation will show the effects of changes in the federal funds rate but will not allow for a causal interpretation, as changes in the interest rates are endogenous to the state of the economy. To obtain exogenous variation in Δr_t, we will use the structural policy shocks introduced by Romer and Romer (2004). The Romer-Romer shocks are the component of policy changes that are orthogonal to the Federal Reserve's information set, Federal Reserve Greenbook projections for GDP, inflation, and unemployment. Taking account of the delay in the publication of the Greenbook, the data are currently available for the period from 1969 to 2015. More specifically, let Δz_t denote the surprise component. We will estimate the following set of local projections using instrumental variables:

$$(2) \qquad y_{t+h} = \alpha_h + \Delta \hat{r}_t \beta_h + x_t \gamma_h + \nu_{t+h}; \text{ for } h = 0, \ldots, H-1,$$

The estimates of $\Delta \hat{r}_t$ come from the first-stage regression:

$$(3) \qquad \Delta r_t = \Delta z_t b + x_t g + \varepsilon_t.$$

Data for the outcome variables and the controls are all standard, publicly available macroeconomic time series. Specific variable definitions and sources are shown in table 2.

IV.B. Effects of Monetary Policy: Results

Our estimates of the response of financial and labor market outcomes to a 100 basis point expansionary monetary policy shock are shown in figure 8. The results use the Romer-Romer shock series as an instrument for changes in the federal funds rate. The estimations show a large response of stock prices that peaks at about 5 percent in less than three years. The effect declines to about 3 percent by year 5 but remains sizeable over the entire horizon. By contrast, the house price response takes more than a year to get started and peaks at a little more than 2 percent after five years. Treasury yields fall on impact but then return to their original level after

Table 2. Macroeconomic Data

Variable	*Description*	*Time Period*	*Source*
Federal funds rate	Federal funds target	11/1988 to 9/2017	FRB
Unemployment rate	Seasonally adjusted unemployment	1/1960 to 9/2017	BLS
Unemployment gap	Difference in Black and white unemployment rates	1/1972 to 9/2017	BLS
Hourly wages	Black and white workers	1/1982 to 9/2017	BLS
Weekly earnings	Black and white workers	1/1982 to 9/2017	BLS
Industrial production	Industrial production index	1/1960 to 9/2017	FRB
Stock price	S&P 500 price	1/1960 to 9/2017	S&P
Inflation	CPI, all urban consumers	1/1960 to 9/2017	BLS
M2 growth	Real money stock	1/1960 to 9/2017	FRB
House price	Case-Shiller house price index	1/1975 to 9/2017	S&P CoreLogic
Dividends	Real dividends, S&P 500	1/1960 to 9/2017	R. Shiller
Corporate bond yield	Moody's seasoned corporate BAA yield	1/1960 to 9/2017	FRB
Treasury yield	Ten-year constant maturity T-note yield	1/1960 to 9/2017	FRB

Sources: Federal Reserve Bank; Bureau of Labor Statistics; S&P CoreLogic Case-Shiller Home Price Indices; R. Shiller, http://www.econ.yale.edu/~shiller/data.htm.

about three years.[17] The coefficient estimates at projection horizons ranging from impact to five years are shown in table 3.[18]

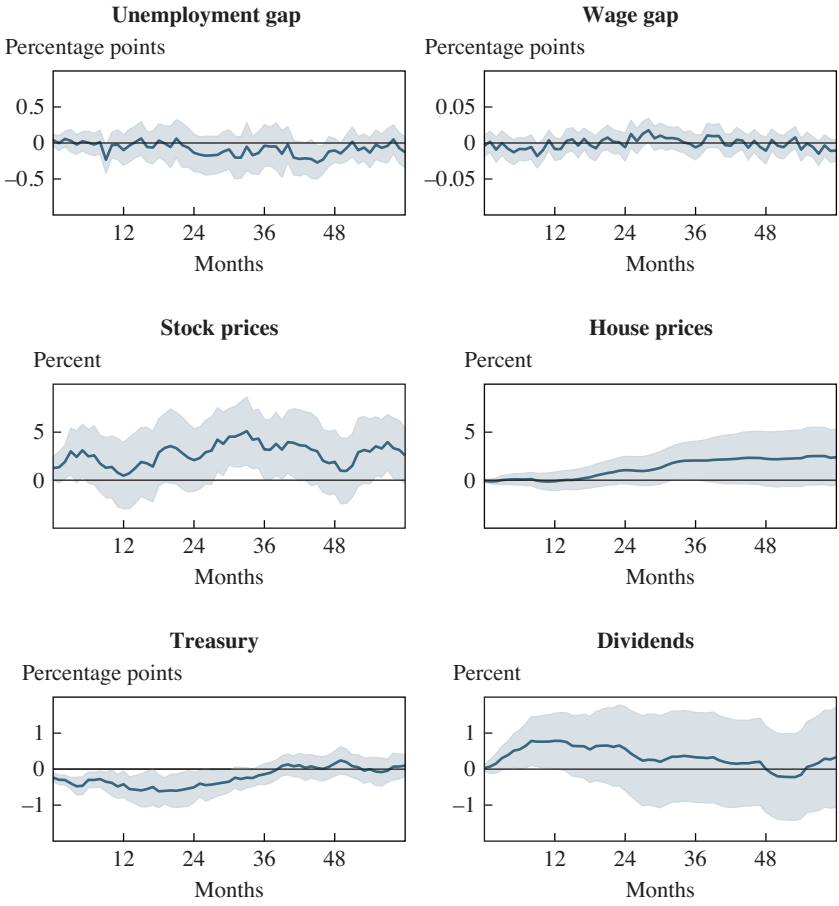
LABOR MARKET OUTCOMES Both the results with the instrumented and uninstrumented change in the federal funds rate indicate that there is a small effect on the unemployment gap, which is sometimes significant at the 90 percent level. After a 100 basis point expansionary shock, the unemployment gap closes by 0.14 percentage points. Similar results are reported in Carpenter and Rodgers (2004), who find that a one-standard-deviation monetary policy shock reduces the Black unemployment rate on average by 0.15 percentage points more than the white unemployment rate. Their estimated effect is also persistent; it declines slightly over time but remains significant even after four years.

The results above do not suggest any discernible effect of an expansionary monetary policy shock on the mean Black-white hourly wage gap. We also examined alternative measures of earnings, namely, the gap in

17. The results can be compared to the simple LP-OLS estimation from equation (1) where the change in the federal funds rate is not instrumented. These results, shown in online appendix figure A.5, are similar. Unsurprisingly, the equity and house price effects are smaller.

18. The effects of the policy shock on inflation and the BAA corporate bond yield are not shown to conserve space.

Figure 8. Effects of a 100 Basis Point Monetary Policy Shock



Source: Authors' calculations.

Note: The figure shows the impulse responses after a Romer-Romer 100 basis point expansionary monetary policy shock. Impulse responses are shown as solid lines and shaded areas show 90 percent confidence bands.The vertical axes show asset price changes as a percentage for stocks and houses, in basis points for ten-year Treasury yields, and in percentage points for the racial unemployment gap.

Table 3. Estimates for Response to 100 Basis Point Expansionary Monetary Policy Shock

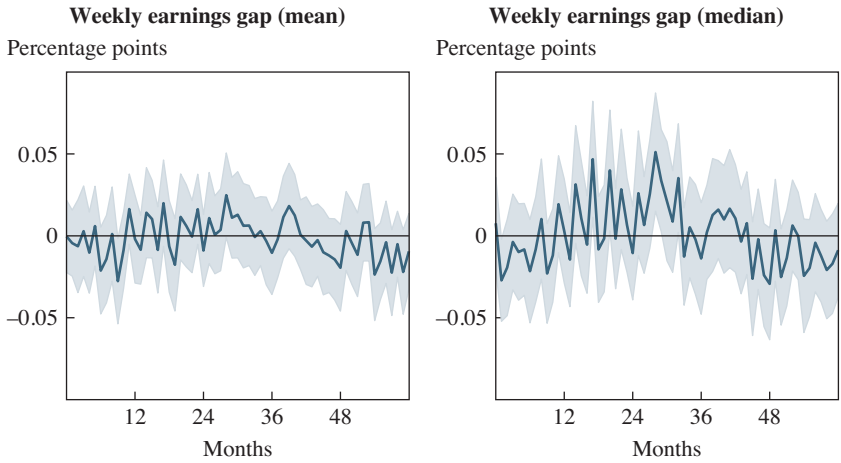
Horizon	*Unemployment gap (%)*	*Wage gap (%)*	*Stock prices*	*House prices*	*Ten-year Treasury yields*	*Dividends (%)*
Zero months	0.038	–0.004	1.268*	–0.074	–0.236***	0.026
	(0.121, –0.045)	(0.011, –0.018)	(2.461, 0.074)	(0.069, –0.217)	(–0.117, –0.355)	(0.119, –0.067)
6 months	0.004	–0.008	2.479	0.080	–0.299**	0.548*
	(0.161, –0.154)	(0.008, –0.024)	(5.254, –0.295)	(0.687, –0.527)	(–0.082, –0.516)	(1.088, 0.007)
12 months	–0.009	–0.008	0.463	–0.118	–0.420*	0.787*
	(0.111, –0.308)	(0.007, –0.024)	(3.903, –2.976)	(0.835, –1.070)	(–0.055, –0.785)	(1.553, 0.020)
24 months	–0.137	–0.006	2.089	1.046	–0.505**	0.566
	(0.123, –0.396)	(0.013, –0.025)	(5.346, –1.169)	(2.475, –0.383)	(–0.085, –0.925)	(1.731, –0.598)
36 months	–0.038	–0.006	3.206*	2.047*	–0.143	0.328
	(0.253, –0.328)	(0.011, –0.023)	(6.333, 0.078)	(4.044, 0.051)	(0.148, –0.434)	(1.575, –0.918)
48 months	–0.104	–0.011	1.905	2.230	0.154	0.001
	(0.061, –0.269)	(0.008, –0.029)	(5.211, –1.401)	(5.049, –0.589)	(0.500, –0.192)	(1.195, –1.193)
60 months	–0.129	–0.011	2.564	2.383	0.097	0.334
	(0.084, –0.342)	(0.004, –0.025)	(5.493, –0.365)	(5.299, –0.533)	(0.393, –0.199)	(1.723, –1.054)

Source: Authors' calculations.

Note: Parentheses below the point estimates at each horizon show the 90 percent confidence interval.

* significance at the 10 percent level; ** significance at the 5 percent level; *** significance at the 1 percent level.

Figure 9. Effects of a 100 Basis Point Monetary Policy Shock on the Weekly Earnings Gap



Source: Authors' calculations.
Note: Impulse responses are shown as solid lines and shaded areas show 90 percent confidence bands.

mean and median weekly earnings of Black and white workers from the BLS. The effect of the policy shock on each of these series is shown in figure 9. The results confirm the previous picture based on average hourly wages, with no discernible effect on the relative weekly earnings. The estimation results for the earnings gap also suggest that hours react little so that employment changes stem mainly from the extensive margin of employment. Our conclusion is that any effect that more accommodative monetary policy has on labor market outcomes of Black Americans is likely to come from employment gains and less from the relative wage effect.

ALTERNATIVE SHOCK SERIES Our estimates above rely on the widely used shock series by Romer and Romer (2004) to instrument the change in the federal funds rate. The Romer-Romer policy shocks are the component of the change in the federal funds rate that is not explained by the Federal Reserve's information set. There are many other ways to estimate policy shocks, including estimates that utilize information from the federal funds futures markets. However, Ramey (2016) shows that estimates of the effects of policy shocks are often sensitive to small changes in technique, definition, or estimation period. Moreover, confidence intervals for policy effects are often wide. Thus, we do not claim to have identified precise point estimates for policy effects, but we suggest that our benchmark estimates with the Romer-Romer shocks are within a plausible range suggested by different approaches.

In online appendix II we show that results with three other shock series are broadly similar to the results shown above. The first series shown is the measure introduced by Bernanke and Kuttner (2005) that sparked interest in the effect of monetary policy on asset prices. It is based on the difference between the federal funds target rate and the rate implied by futures contracts. The second series is based on shocks from Gertler and Karadi (2015) that use high-frequency responses from the federal funds futures markets immediately following each Federal Open Market Committee meeting to identify a policy shock. The third series uses monetary policy surprises based on federal funds futures contract from Paul (2020).[19]

V. Earnings and Portfolio Effects of Monetary Policy

The empirical results in section IV show substantial and persistent positive asset price effects of a surprise monetary easing, in combination with a reduction in the Black-white unemployment gap. In this section, we use these estimates to calculate the effects of a monetary policy shock on the wealth of the average Black and white households. Since the wealth distribution is highly skewed, we also examine the portfolio effects along the wealth distribution and around the median. Finally, we calculate the effect of a monetary policy shock on the gap between Black and white earnings and compare the size of the portfolio and earnings effects over different horizons.

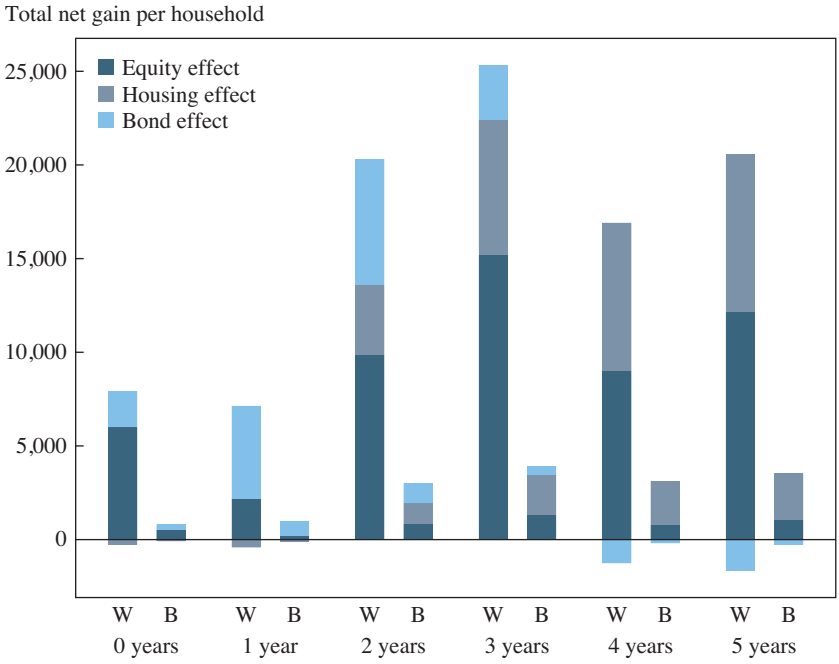
V.A. Effects on Household Wealth

One additional step is needed before we can estimate the impact of a monetary policy shock on wealth. For bonds, we need to transform the effect on interest rates into a change in the asset price using an assumption about duration. We use duration estimates taken from Bloomberg on October 30, 2020. The average duration of outstanding ten-year Treasuries (9.47), mortgage-backed securities (5.43), and corporate bonds (7.07) are applied to the corresponding asset categories in the SCF data.[20]

19. We also estimated a time-varying VAR (TV-VAR), following Paul (2020). The TV-VAR aims to capture different responses of asset prices to monetary policy shocks over time, depending, for instance, on risk appetite in markets. The results are generally similar and available upon request. In particular, the stock price response is very persistent also with the TV-VAR, and even larger than in the LP-IV regressions.

20. We use corporate bond duration and yield for corporate and foreign bonds, Treasury duration and yield for government, state, and municipal bonds, and mortgage-backed securities duration and corporate yield for mortgage-backed bonds.

Figure 10. Capital Gains for Black and White Households from Monetary Policy Shocks over Time



Source: Authors' calculations.

Note: The figure shows the average wealth effects for Black (B) and white (W) households after a 100 basis point monetary shock over time. The wealth effects are computed by combining the estimates from table 3 with portfolio data from the SCF.

To be consistent with stock and house price changes, which are real, the nominal change in each bond wealth category is deflated using the estimated responses of inflation to the policy shock.

We are now in a position to estimate the effects of the monetary policy shock (a 100 basis point surprise decline in the federal funds rate) on household wealth. The portfolio capital gains on each asset class are shown in figure 10. At every horizon, the total capital gains from an unanticipated monetary policy accommodation are much larger for white households than the gains to Black households. The largest effects are after three years, reaching $25,300 for white households and $3,900 for Black households. The biggest impact comes from the large and persistent effect of equity prices. The house price effect increases for around three years and then remains roughly constant. The bond effects are small because bond holdings are only a small fraction of total wealth for both Black and white households.

An unanticipated monetary policy accommodation leads to asset price changes that benefit white households to a much larger extent than Black households because average white wealth is much larger and a larger fraction is held in equities, where asset prices react most strongly.

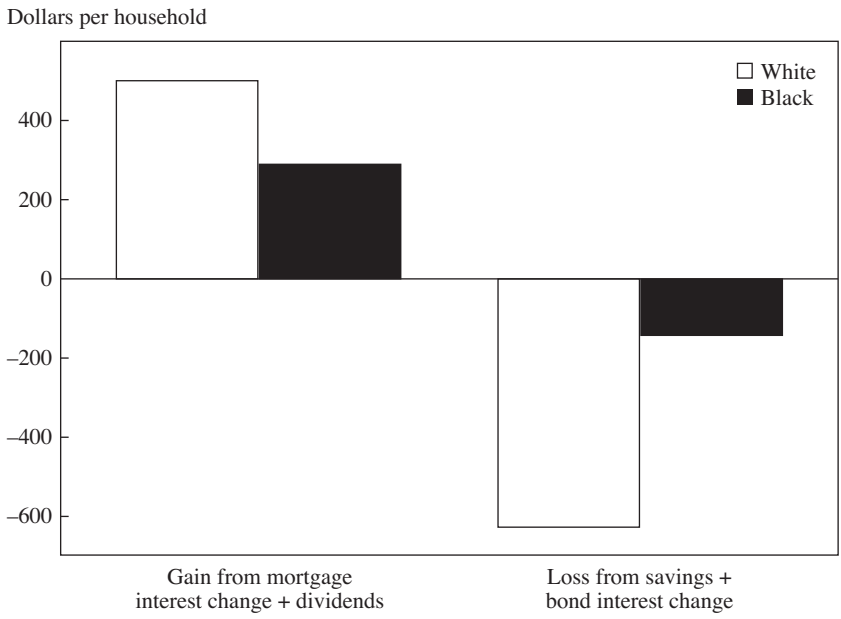
In addition to the direct effects on capital gains from the monetary shock, there are also indirect effects on capital income. That is, monetary policy shocks can reduce mortgage interest rates and the interest earned on deposit-type assets and corporate and mortgage-backed bonds and increase dividends. We estimate the effects based on the results from table 3 and the method described in section III.B. We assume that the impact of the 100 basis point accommodative monetary shock on mortgage rates is given by the impact on the ten-year Treasury rate and use the estimate at a one-year horizon to calculate the savings on annual mortgage payments. For liquid assets, we assume that the decline in interest earnings is the same size as the monetary policy shock, 100 basis points. For bonds, we use the effects on Treasury (–42 basis points) and BAA corporate bond yields (–36 basis points) after one year, and for dividends we apply the percentage change after one year (0.8 percent) to average Black and white dividend income from the SCF.

In figure 11, we show the capital income (dividend and interest rate) effects from the accommodative monetary shock. Black households, with small deposit balances to begin with, lose little from lower interest rates, and on net, the average Black household gains more from mortgage refinancing. White household deposit interest losses, which amount to around $600, are about $100 larger than the average annual gains from refinancing and dividend increases. This calculation is again based on a scenario in which all households refinance. Lower refinancing rates of Black households would increase the gap between Black and white households (Gerardi, Willen, and Zhang 2021).

ON THE PERSISTENT EFFECTS OF POLICY SHOCKS Although monetary policy shocks by construction capture cyclical variation, they can still have persistent effects on inequality. First, we find that asset prices change after monetary policy shocks for an extended period of five years. Our results build on a growing body of literature that estimates persistent asset price changes in response to monetary policy shocks (Bernanke and Kuttner 2005; Jordà, Schularick, and Taylor 2015; Paul 2020; Rigobon and Sack 2004). Such a period can easily account for 10 percent of the economically active lifetime of a household.

Second, recent theoretical and empirical work suggests that monetary policy shocks can affect the long-run equilibrium interest rate (Bianchi,

Figure 11. Effects of Monetary Policy Shocks on Capital Income for Black and White Households after One Year



Source: Authors' calculations.

Lettau, and Ludvigson 2022; Rungcharoenkitkul, Borio, and Disyatat 2021). Bianchi, Lettau, and Ludvigson (2022) show that monetary policy leads to regime shifts with long-lasting effects on relative asset prices. In this case, there can be permanent impacts on asset prices.

Moreover, distributional effects may persist even if gains and losses average out over time and asset prices revert to an equilibrium, as indicated in theory (Auclert 2019). This is because portfolio decisions by households are often driven by changes in their life cycle situation rather than financial returns. For example, household formation or changes in family composition can lead to portfolio adjustments such as the purchase or sale of a house. In such instances, households cannot simply wait for asset prices to revert back to their long-run level without welfare consequences from not adjusting their asset positions. In general, the life cycle puts young households systematically on the buyer side and older households on the seller side of the market and will induce constant trading needs that are not governed by asset price movements. That is, capital gains are often realized by households due to life cycle events such as marriage, divorce, family

formation, job loss, or job change. Hence, differences along racial lines in household demographic structure or unemployment experience can induce differences in the propensity to buy and sell assets, in addition to the racial differences in the exposure to asset price change.

Also, asset price changes may alleviate or tighten collateral constraints as, for example, discussed in Iacoviello (2005). An expansionary monetary policy shock relaxes borrowing constraints and offers the opportunity to access additional credit for consumption or investment. This collateral effect will likely play out differently along racial lines, as the fraction of homeowners is larger among white households and housing is the key asset through which the collateral channel can work. Even a short-lived price change can trigger this channel, given that borrowing constraints only have to hold when the loan is originated.

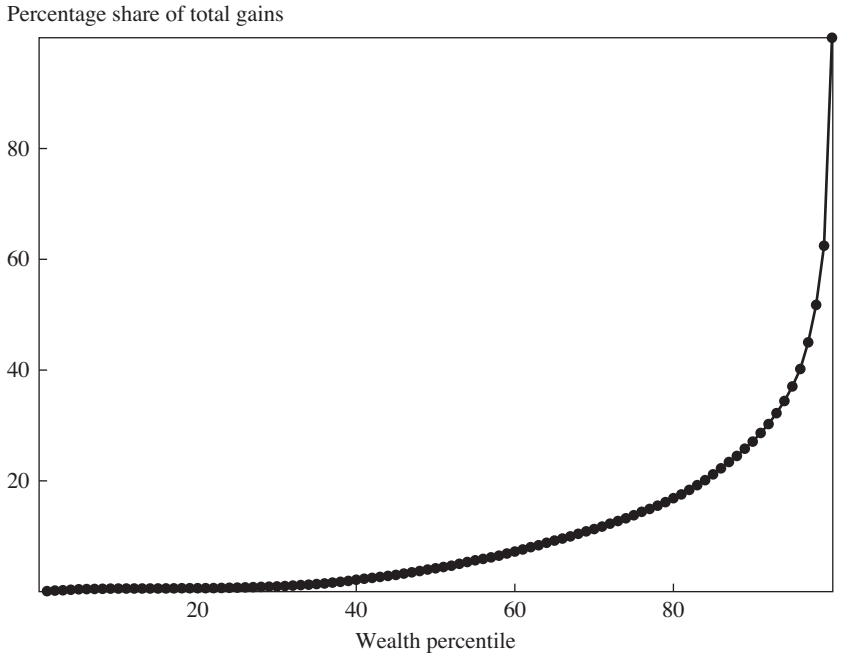
V.B. Portfolio Effects along the Wealth Distribution

Our estimates of the portfolio effects of asset price changes shown above consider the average Black and white household. Since the US wealth distribution is highly skewed (Kuhn and Ríos-Rull 2016; table 1), these results might not be fully representative. In this section, we examine this issue in two ways. First, we show the distribution of gains from a monetary policy surprise along the wealth distribution, and second, we look at results that restrict the sample to households around the median.

DISTRIBUTION OF GAINS The distributional implications of the portfolio effects after five years are shown in figure 12. The figure is a Lorenz curve of the wealth gains from a 100 basis point monetary policy surprise along the wealth distribution for all households. About 75 percent of all gains accrue to households in the top 10 percent of the wealth distribution and about 38 percent go to the top 1 percent. Notably, this distribution is substantially more unequal than the distribution of wealth itself. The facts that equity gains account for a large share of the total gains and equity holdings are highly concentrated along the wealth distribution lead to a high concentration of the gains from monetary policy in the—mainly white—top 10 percent of the wealth distribution.[21]

21. The Black households in our data are very unequally distributed along the wealth distribution. Among the bottom 50 percent of households, the share of Black households is 24 percent. Their share is 9 percent among households between the 50th and 90th percentile. Only 2 percent of households among the top 10 percent wealthiest households are Black.

Figure 12. Lorenz Curve of Estimated Portfolio Gains after Expansionary Monetary Policy Shock



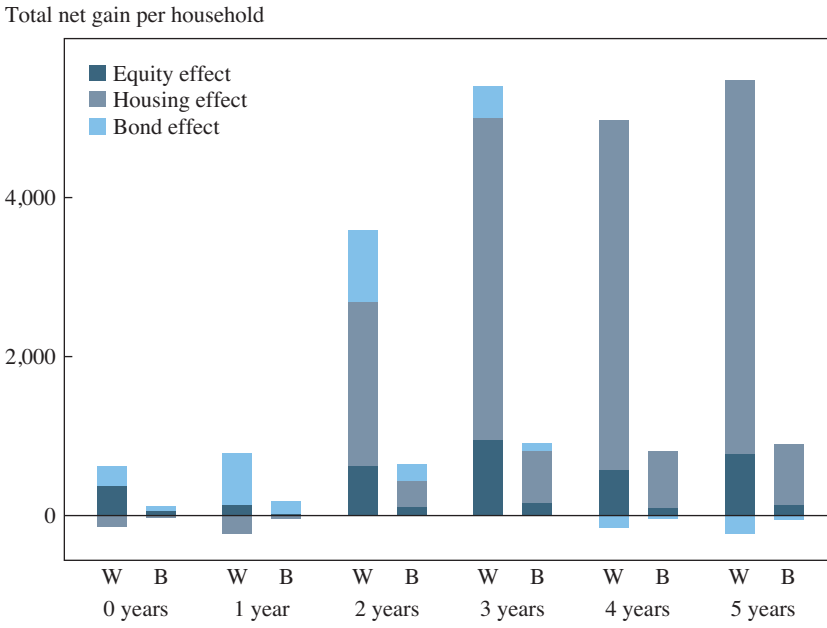
Source: Authors' calculations.

Note: The graph shows the Lorenz curve of the total portfolio effect in year 5 after an expansionary 100 basis point monetary policy shock.

HOUSEHOLDS AROUND THE MEDIAN Since portfolio gains are so highly concentrated among wealthy households, it is reasonable to suspect that the wealth gap among more "typical" households is less affected by asset price changes. In order to examine this, we look at Black and white households around the median, which we define as households between the 40th and 60th percentiles of their respective wealth distributions.

The portfolio effects of a monetary policy surprise on Black and white households around the median are shown in figure 13. Comparing the effects around the median to the average effects in figure 10, we find that gains are smaller in levels but that the relative differences between Black and white households persist. We still find that the gains of white households are more than four times larger than for Black households. The gains around the median differ in their composition relative to the mean effects because of the differences in the portfolio composition along the wealth distribution.

Figure 13. Capital Gains for Black and White Households around the Median from Monetary Policy Shocks over Time



Source: Authors' calculations.

Note: The figure shows the average wealth effects for Black (B) and white (W) households around the median after a 100 basis point monetary shock over time. The wealth effects are computed by combining the estimates from table 3 with portfolio data from the SCF. The underlying portfolios are constructed by averaging across all households between the 40th and 60th percentile of the respective wealth distributions separately for Black and white households.

We find that around the median, most of the gains stem from housing, whereas equity gains are the largest part at the mean. As a result, it takes about two years for gains to accumulate, and they are persistent after that. The capital gains are about the same size in year 5 as in year 3.

In table 1, we reported that a large share of Black households do not own any assets of several types and if they do, their holdings are often small. To see the implications of this, we look at the shares of Black and white households who have portfolio gains that are less than 1 percent of their annual income five years after an expansionary shock. We refer to households with a portfolio gain below 1 percent of income as having no portfolio effect. We find that about one-fourth of white households (24 percent) have no portfolio effect after five years. By contrast, the share among Black households is more than twice as large (53 percent). Hence, almost half of Black

households are left with no portfolio gains five years after an expansionary monetary policy shock.[22]

By construction, Black and white households with similar portfolios will have similar capital gains. Figure 14, panel a, shows that capital gains for Black and white households are indeed similar when looking at households between the 40th and 60th percentile of the *overall* wealth distribution. The effects are only slightly smaller for Black households, mostly due to somewhat smaller housing capital gains. The remaining differences in capital gains disappear when normalizing by income, as shown in online appendix figure A.3. However, Black households are underrepresented in the middle and upper parts of the aggregate wealth distribution. Online appendix figure A.4 shows that the share of Black households in the upper half of the wealth distribution has consistently been lower than the overall population share of Black households since the 1950s. Moreover, since the 1970s Black households have become less likely to make it to the top 10 percent, and more likely to be in the bottom 50 percent.

Figure 14, panel b, looks at capital gains for Black and white households around the median of the aggregate *income* distribution. Here, we again see pronounced differences between Black and white households. In other words, even Black households who by construction have similar incomes as white households do have lower wealth and therefore reap lower capital gains after expansionary monetary policy shocks. On average, the capital gains of white households between the 40th and 60th percentile of the aggregate income distribution are two to three times larger than those of their Black counterparts.

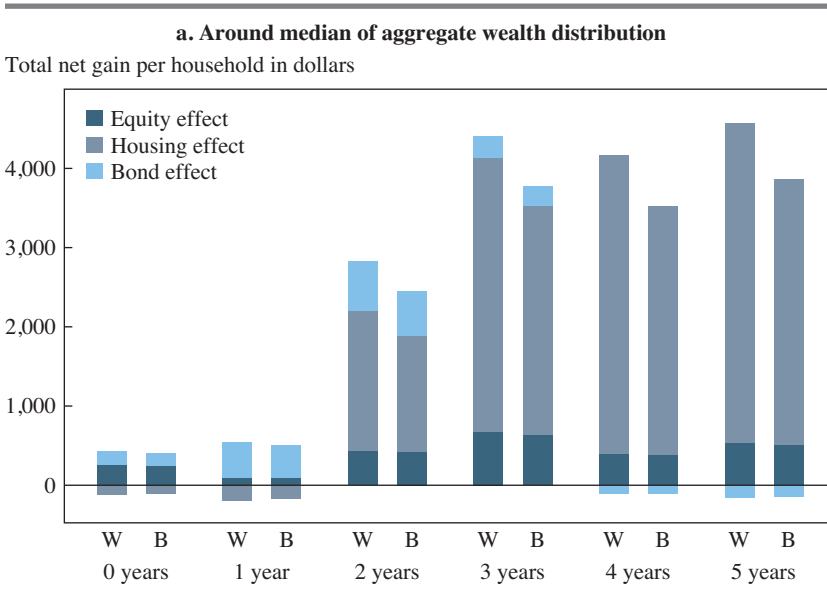
In figure 15, we show the capital income effects of a monetary policy shock for Black and white households around the median, using the same assumptions as in the results for the mean households in figure 11. White households around the median have gains from mortgage refinancing which are about three times larger than the gains for Black households, a much larger difference than for the mean household because many more of the Black households around the median do not own a home.

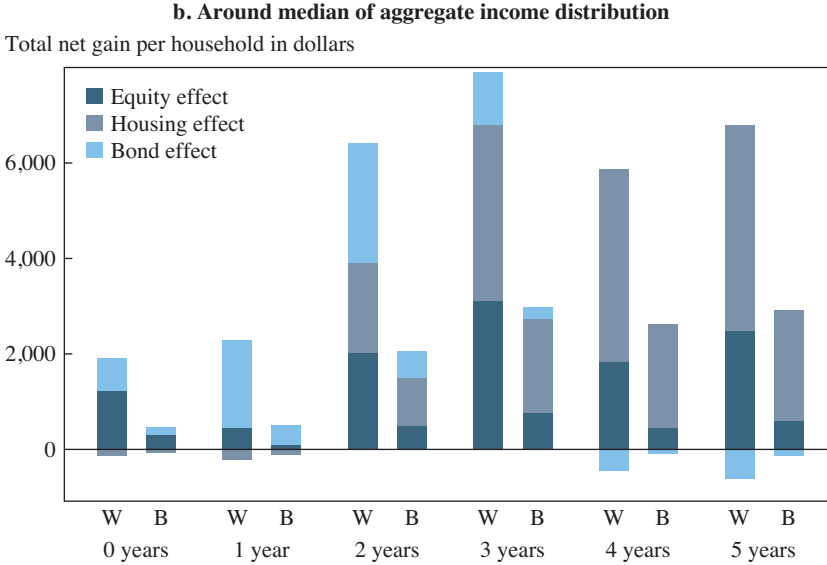
V.C. Differences among Households by Marital Status and Sex

The analysis so far has examined the portfolio effects for Black and white households and has not addressed any additional demographic characteristics, although there are significant differences in the demographic

22. If we consider a 5 percent threshold instead of the 1 percent threshold, the shares increase to 41 percent for white households and 68 percent for Black households.

Figure 14. Capital Gains for Black and White Households around the Median of the Aggregate Wealth and Income Distribution

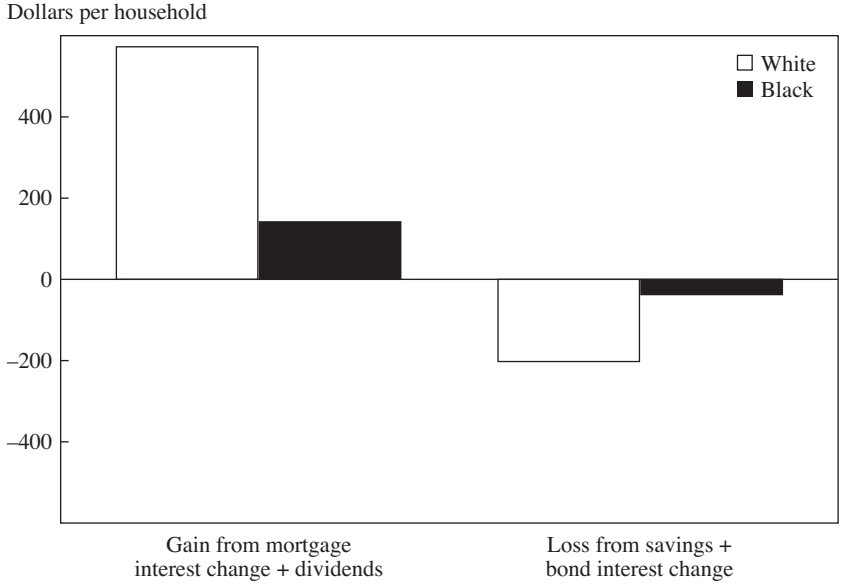




Source: Authors' calculations.

Note: The figure shows the average wealth effects for Black and white households around the median of the *aggregate* wealth (panel a) and income (panel b) distribution after a 100 basis point monetary shock over time. The wealth effects are computed by combining the estimates from table 3 with portfolio data from the SCF. The underlying portfolios are constructed by averaging across all households between the 40th and 60th percentile of the aggregate wealth (income) distribution.

Figure 15. Effects of Monetary Policy Shocks on Capital Income for Black and White Households around the Median after One Year



Source: Authors' calculations.

Note: The graph shows the average gains for Black and white households around the median after a decline in mortgage interest rates and increase in dividend income, and their average losses after a decline in savings and bond interest rates, as implied by the monetary policy shock after one year. The underlying portfolios are constructed by averaging across all households between the 40th and 60th percentile of the respective wealth distributions for Black and white households.

composition of households with white and Black heads.[23] Many more white households consist of married or cohabiting couples and more of the single Black households are led by women. In online appendix IV, we show the effects of a Romer-Romer monetary policy shock first on households separated by marital status and second on single households separated by sex of the head of household. The overall findings of the paper are unaffected when we examine results disaggregated by household type. The portfolio gains of white households of any type are almost always larger in both absolute terms and relative terms than for the corresponding group of Black households. Notably, the gains to single white households are typically larger than the gains to married Black households.

23. As only 15 percent of the SCF households have a Black head of household, the granularity of further breakdowns is limited.

V.D. Quantifying the Earnings Effect

Our estimates in section IV.B indicate that an accommodative monetary policy shock reduces the unemployment rate for Black households more than for white households, although there are no discernible effects on the gap in wages. Nevertheless, the employment effects will reduce the gap in mean earnings by increasing the relative number of people receiving labor income. In this section, we aim to quantify the earnings effects from the reduced unemployment rates. We combine the low-frequency 2019 SCF data on labor income with our estimates of the impact of monetary policy shocks on the unemployment gap. Using this estimate, we are in a position to compare the relative gains from the earnings and wealth effects for Black and white households.

For our calculation, we focus on prime-age household heads (age 25–55) and on the information if the head of household has been unemployed during the twelve months before the interview.[24] There are large differences in the unemployment experiences of Black and white households. The share of Black household heads experiencing unemployment in the year prior to the interview is 12.4 percent, while the share for white household heads is 8.3 percent. Comparing earnings of households who have been and who have not been unemployed during the past twelve months, we find that average earnings of Black households whose head has not been unemployed are $56,200. For households whose head experienced unemployment within the last twelve months, the average annual labor income is $27,500.[25] By contrast, we find that white households who experienced unemployment during the last twelve months still report average earnings of $50,300—almost the level of Black households without unemployment experience. White households without unemployment experience over the last twelve months report an average labor income of more than $103,000 in the 2019 SCF data.

To derive the earnings effect, we multiply the difference in earnings between Black households that have and have not experienced unemployment by our estimates of the impact of monetary policy on the differential between Black and white unemployment rates from table 3. We then make a conservative assumption in order to relate the change in the unemployment gap to earnings changes. In particular, we assume that each household

24. We consider the last twelve months rather than the current labor force status at the interview because the surveyed labor income also refers to the previous calendar year.

25. Sample sizes are small: we observed 182 white households and 64 Black households whose head of household reported unemployment during the last twelve months.

that finds employment receives the average earnings gain of a Black household with a household head who did not experience unemployment, thus the earnings gain is \$56,200 – \$27,500 = \$28,700. The relative income gain of Black households is computed by multiplying the estimated impact of the monetary policy shock on the unemployment gap with the average earnings gain of \$28,700.

More formally, let us denote the estimated effect on the unemployment gap at projection horizon h by $\Delta_h u$ and the earnings gain from leaving unemployment for Black households by $\Delta Y^B = Y_E^B - Y_U^B$ where Y_E^B denotes average labor income for Black households who have not been unemployed over the past twelve months and Y_U^B denotes average labor income of Black households who have been unemployed at least for some time in the past year. In the 2019 SCF data, we find ΔY^B = \$28,700. Our estimate for the relative earnings gain for Black households relative to white households in period h after the shock, $\Delta_h Y$, is thus

$$\Delta_h Y = \Delta_h u \Delta Y^B = \Delta_h u \left(Y_E^B - Y_U^B \right).$$

The effect on the unemployment gap in table 3 peaks after two years, when the unemployment rate gap is reduced by 0.137 percentage points. The relative earnings gain is found by multiplying this number with the average earnings gain, which yields a relative gain per Black household of \$39.30, or 0.07 percent of annual total income for all Black households.

V.E. Comparing Earnings and Portfolio Effects

The impact of the monetary policy shock on the *difference* in earnings between Black and white households is $\Delta_h Y$. The appropriate comparison is to the *difference* in capital gains accruing to Black and white households. The calculation above showed that the maximum earnings difference occurs after two years, $\Delta_2 Y$ = \$39.30. At the same horizon, capital gains received by the average white household are about \$17,300 more than those of the average Black household (see figure 10). This comparison suggests that the relative portfolio gain for white households is orders of magnitude larger than the relative earnings gain for Black households.

However, there are important conceptual differences between the capital gains on assets and changes in earnings. First, the earnings effect applies to the flow of earnings, while the capital gains are a gain on the stock of wealth. Thus, the capital gain is a onetime change in the valuation of assets, while the earnings effect applies to incomes year by year. To take this into account, we compare the difference in capital gains between white and Black households over the five-year horizon to the accumulated estimate

of the differential earnings effect over this time period. Second, the earnings effect directly increases the consumption possibilities of households, whereas the capital gain needs to be realized first before it increases households' consumption possibilities.

The year-by-year accumulated earnings effects and the difference in the portfolio effects for Black and white households as a percentage of each group's income were shown in figure 1. As pointed out earlier, the earnings effects are tiny in comparison to the portfolio effects.

Finally, we turn to the consumption effects of capital gains in order to make the portfolio and earnings effects more directly comparable. We find that under plausible and conservative assumptions, the consumption consequences of the wealth effect for white households is larger than the entire earnings effect for Black households. There is a large body of literature that estimates the marginal propensity to consume (MPC) out of wealth. In recent work, Chodorow-Reich, Nenov, and Simsek (2021) exploit regional variation in stock market exposure in the United States and estimate an annual MPC out of capital gains of 3.2 percent.[26] As this estimated consumption response to capital gains is annual, it implies a consumption increase in each year with capital gains.

Our estimated capital gain five years after an accommodative monetary policy shock is about $15,600 more for white than for Black households (figure 10), which corresponds to additional consumption expenditures of almost $500 in year 5 alone, abstracting from consumption increases in other years, making this a conservative estimate for the differential consumption response. Thus the portfolio effect on consumption for white households in year 5 after the shock is already three and a half times as large as the entire accumulated relative earnings effect for Black households of $134.

Given that we are unaware of estimates for the MPC out of capital gains by race, we assume that the MPCs are identical for Black and white households. If Black households had a higher MPC out of capital gains, the differential capital gains would have to be adjusted for these differences. However, if we assume that the MPC estimate of 3.2 estimated by Chodorow-Reich, Nenov, and Simsek (2021) is a population-weighted average between Black and white households' MPCs, we can calculate how large Black and white MPCs would have to be in order to offset the relative

26. This estimate is typical of the literature on the MPC out of capital gains. Poterba (2000) and Paiella (2009) summarize the literature. More recently Di Maggio, Kermani, and Majlesi (2020) and Chodorow-Reich, Nenov, and Simsek (2021) present estimates based on micro data.

income effect of $134. This back-of-the-envelope calculation shows that Black MPCs would have to be roughly three times larger than white MPCs for the relative consumption effect from capital gains to be as large as the relative earnings effect. Ganong and others (2020) estimate that the consumption response of Black households to typical labor income shocks is about 50 percent larger than for white households. While this paper does not estimate MPCs out of capital gains, it may still serve as a guideline for what plausible differences in Black and white MPCs might look like. We are therefore confident that our result of a larger relative consumption gain for white households compared to the relative earnings gain for Black households remains valid under plausible assumptions on Black and white MPCs out of capital gains, even if they were not equal.

There is evidence that expansionary monetary policy improves the labor market situation of Black households more than for white households. Yet, when we contrast the consumption effects of capital gains from asset price changes to the earnings effect, we find that the earnings gains of Black households are dwarfed by the consumption changes implied by the portfolio gains of white households.[27]

VI. Conclusion

We have shown that policy shocks that change asset prices have differential effects on the wealth of Black and white households. White households gain more because they have more financial wealth and hold portfolios that are more concentrated in interest rate–sensitive assets such as equities. At the same time, monetary policy shocks reduce the gap between Black and white unemployment rates and entail larger earnings gains for Black households. Bringing the two together, however, leads to a stark finding: the reduction in the earnings gap pales in comparison to the effects on the wealth gap.

Our analysis therefore does not bode well for the suggestion made by politicians and central bankers that a more accommodative monetary policy helps alleviate racial inequalities. With the instruments available—all of which work through effects on asset prices and interest rates—a central bank would not be able to design policies for an income gap reduction objective without increasing wealth inequality. Clearly, this does not mean

27. The earnings effects for single households led by men and women are shown in online appendix figure A.12. They are small when compared to the corresponding portfolio effects.

that achieving racial equity should not be a first-order objective for economic policy. We strongly think it should. But the tools available to central banks might not be the right ones and could possibly be counterproductive.

One possible conclusion of our research is that there is no role for central banks in addressing society's concern with racial gaps or inequality. This conclusion would be consistent with the traditional view that a central bank should have a singular focus on price stability. However, as Wachtel and Blejer (2020) show, this idealized view of central banking does not agree with historical experience and has been repeatedly challenged by the financial and pandemic crises. Furthermore, the traditional view conflicts with the concerns expressed by Federal Reserve officials about racial gaps.[28]

In light of our results, is there any role for central banks in addressing the challenges of racial wealth and income gaps? No Federal Reserve official has suggested that the conventional tools of monetary policy should be used to address racial gaps except as an offshoot of the full employment mandate. They do imply that the Federal Reserve has a role in bringing public attention to wealth and income gaps in a way that might influence both fiscal policy and the behavior of the private sector. More concretely, the Federal Reserve System has become an influential center for research on inequality.[29] Even these efforts have not been without criticism; Senator Pat Toomey, the ranking member of the Senate Banking Committee, tried to launch a review of Federal Reserve mission creep and "research [that] appears to be focused on how matters unrelated to monetary policy impact narrow subgroups of people."[30] So an additional insight from our research is that despite all the rhetoric about central bank independence, central bankers are inevitably part of the political and policy discourse about inequality.

28. See, for example, Powell (2020), Bostic (2020), and Daly (2020).

29. For example, the Minneapolis Federal Reserve established the Opportunity and Inclusive Growth Institute in 2017, and the St. Louis Federal Reserve established the Institute for Economic Equity in 2021.

30. United States Senate Committee on Banking, Housing, and Urban Affairs, "Toomey Launches Review of Mission Creep by Regional Federal Reserve Banks," press release, March 29, 2021, https://www.banking.senate.gov/newsroom/minority/toomey-launches-review-of-mission-creep-by-regional-federal-reserve-banks.

ACKNOWLEDGMENTS We are grateful to Klaus Adam, Adrien Auclert, James Bullard, Ellora Derenoncourt, William Darity, Keshav Dogra, Michael Fleming, Linda Goldberg, Jordi Galí, Bev Hirtle, Ethan Ilzetzky, Chi Hyun Kim, Anna Kovner, Alexander Kriwoluzky, Keith Kuester, Òscar Jordà, David Lucca, Ben Moll, Emanuel Mönch, Gernot Müller, Marco del Negro, Emil Holst Partsch, Pascal Paul, Anna Rogantini Picco, Andrea Tambalotti, Alan Taylor, and John Williams for comments and suggestions. The views expressed herein are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of New York, the Federal Reserve System, or Danmarks Nationalbank. Dean Parker, Sven Eis, Francisco Ruela, and Siddhi Jagdale provided excellent research assistance. All errors are our own. The paper benefited from a visit by Paul Wachtel as a fellow at the Frankfurt Center for Advanced Studies: Foundations of Law and Finance funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG)—project FOR 2774.

References

Adam, Klaus, and Panagiota Tzamourani. 2016. "Distributional Consequences of Asset Price Inflation in the Euro Area." *European Economic Review* 89:172–92.

Aladangady, Aditya, and Akila Forde. 2021. "Wealth Inequality and the Racial Wealth Gap." FEDS Notes. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/econres/notes/feds-notes/wealth-inequality-and-the-racial-wealth-gap-20211022.htm.

Albert, Juan-Francisco, and Nerea Gómez-Fernández. 2021. "Monetary Policy and the Redistribution of Net Worth in the U.S." *Journal of Economic Policy Reform*. doi.org/10.1080/17487870.2021.1895778.

Aliprantis, Dionissi, and Daniel R. Carroll. 2019. "What Is Behind the Persistence of the Racial Wealth Gap?" Working Paper. Federal Reserve Bank of Cleveland. https://www.clevelandfed.org/newsroom-and-events/publications/economic-commentary/2019-economic-commentaries/ec-201903-what-is-behind-the-persistence-of-the-racial-wealth-gap.aspx.

Amberg, Niklas, Thomas Jansson, Mathias Klein, and Anna Rogantini Picco. Forthcoming. "Five Facts about the Distributional Income Effects of Monetary Policy." *American Economic Review: Insights*.

Andersen, Asger Lau, Niels Johannesen, Mia Jørgensen, and José-Luis Peydró. 2021. "Monetary Policy and Inequality." Working Paper. https://econ-papers.upf.edu/papers/1761.pdf.

Auclert, Adrien. 2019. "Monetary Policy and the Redistribution Channel." *American Economic Review* 109, no. 6: 2333–67.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2020. "Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model." Working Paper. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26647.

Bartscher, Alina K., Moritz Kuhn, Moritz Schularick, and Ulrike I. Steins. 2021. "Household Debt and Inequality in the U.S., 1950–2019." Working Paper.

Bayer, Patrick, and Kerwin Kofi Charles. 2018. "Divergent Paths: A New Perspective on Earnings Differences between Black and White Men since 1940." *Quarterly Journal of Economics* 133, no. 3: 1459–501.

Berger, David, Veronica Guerrieri, Guido Lorenzoni, and Joseph Vavra. 2018. "House Prices and Consumer Spending." *Review of Economic Studies* 85, no. 3: 1502–42.

Bernanke, Ben S. 2020. "The New Tools of Monetary Policy." *American Economic Review* 110, no. 4: 943–83.

Bernanke, Ben S., and Kenneth N. Kuttner. 2005. "What Explains the Stock Market's Reaction to Federal Reserve Policy?" *Journal of Finance* 60, no. 3: 1221–57.

Bhutta, Neil, and Benjamin J. Keys. 2016. "Interest Rates and Equity Extraction during the Housing Boom." *American Economic Review* 106, no. 7: 1742–74.

Bianchi, Francesco, Martin Lettau, and Sydney C. Ludvigson. 2022. "Monetary Policy and Asset Valuation." *Journal of Finance* 77, no. 2: 967–1017.

Boerma, Job, and Loukas Karabarbounis. 2021. "Reparations and Persistent Wealth Gaps." Working Paper 28468. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28468.

Bostic, Raphael. 2020. "A Moral and Economic Imperative to End Racism." *1619: Journal of African American Studies* 1, no. 2 (special edition). https://www2.ccsu.edu/afamjournal/?article=484.

Bricker, Jesse, Lisa J. Dettling, Alice Henriques, Joanne W. Hsu, Lindsay Jacobs, Kevin B. Moore, Sarah Pack, John Sabelhaus, Jeffrey Thompson, and Richard A. Windle. 2017. "Changes in US Family Finances from 2013 to 2016: Evidence from the Survey of Consumer Finances." *Federal Reserve Bulletin* 103:1.

Bricker, Jesse, Alice Henriques, Jacob Krimmel, and John Sabelhaus. 2016. "Measuring Income and Wealth at the Top using Administrative and Survey Data." *Brookings Papers on Economic Activity*, Spring, 261–331.

Brunnermeier, Markus, and Yuliy Sannikov. 2012. "Redistributive Monetary Policy." In *Economic Policy Symposium Proceedings: The Changing Policy Landscape.* Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Caramp, Nicolas, and Dejanir H. Silva. 2021. "Monetary Policy and Wealth Effects: The Role of Risk and Heterogeneity." Working Paper. https://www.nicolascaramp.com/files/wealth-effects-risk.pdf.

Carpenter, Seth B., and William M. Rodgers. 2004. "The Disparate Labor Market Impacts of Monetary Policy." *Journal of Policy Analysis and Management* 23, no. 4: 813–30.

Chetty, Raj, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2020. "Race and Economic Opportunity in the United States: An Intergenerational Perspective." *Quarterly Journal of Economics* 135, no. 2: 711–83.

Chodorow-Reich, Gabriel, Plamen T. Nenov, and Alp Simsek. 2021. "Stock Market Wealth and the Real Economy: A Local Labor Market Approach." *American Economic Review* 111, no. 5: 1613–57.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia. 2017. "Innocent Bystanders? Monetary Policy and Inequality." *Journal of Monetary Economics* 88:70–89.

Colciago, Andrea, Anna Samarina, and Jakob de Haan. 2019. "Central Bank Policies and Income and Wealth Inequality: A Survey." *Journal of Economic Surveys* 33, no. 4: 1199–231.

Daly, Mary C. 2020. "Is the Federal Reserve Contributing to Economic Inequality?" Speech given at the University of California, Irvine Virtual Event, October 13. https://www.frbsf.org/our-district/press/presidents-speeches/mary-c-daly/2020/october/is-the-federal-reserve-contributing-to-economic-inequality/.

Dettling, Lisa J., Joanne W. Hsu, Lindsay Jacobs, Kevin B. Moore, and Jeffrey P. Thompson. 2017. "Recent Trends in Wealth-Holding by Race and Ethnicity: Evidence from the Survey of Consumer Finances." FEDS Notes. Washington: Board of Governors of the Federal Reserve System.

Di Maggio, Marco, Amir Kermani, and Kaveh Majlesi. 2020. "Stock Market Returns and Consumption." *Journal of Finance* 75, no. 6: 3175–219.

Emmons, William. 2020. "Housing Wealth Climbs for Hispanics and Blacks, Yet Racial Wealth Gaps Persist." Blog post, April 1, Federal Reserve Bank of St. Louis On the Economy. https://www.stlouisfed.org/on-the-economy/2020/april/housing-wealth-climbs-hispanics-blacks-racial-wealth-gaps-persist.

Emmons William, Ana H. Kent, and Lowell R. Ricketts. 2019. "Just How Severe Is America's Racial Wealth Gap?" Blog post, October 9, Federal Reserve Bank of St. Louis Open Vault. https://www.stlouisfed.org/open-vault/2019/october/how-severe-americas-racial-wealth-gap.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, and Michael Westberry. 2021. *Integrated Public Use Microdata Series, Current Population Survey: Version 9.0* [data set]. Minneapolis: University of Minnesota.

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. https://www.nber.org/papers/w27552.

Gerardi, Kristopher S., Paul S. Willen, and David Hao Zhang. 2021. "Mortgage Prepayment, Race, and Monetary Policy." Working Paper. Federal Reserve Bank of Boston. https://www.bostonfed.org/publications/research-department-working-paper/2020/mortgage-prepayment-race-and-monetary-policy.aspx.

Gertler, Mark, and Peter Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7, no. 1: 44–76.

Glover, Andrew, Jonathan Heathcote, Dirk Krueger, and José-Víctor Ríos-Rull. 2020. "Intergenerational Redistribution in the Great Recession." *Journal of Political Economy* 128, no. 10: 3730–78.

Greenwald, Daniel L., Matteo Leombroni, Hanno Lustig, and Stijn Van Nieuwerburgh. 2021. "Financial and Total Wealth Inequality with Declining Interest Rates." Working Paper 28613. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28613.

Holm, Martin Blomhoff, Pascal Paul, and Andreas Tischbirek. 2021. "The Transmission of Monetary Policy under the Microscope." *Journal of Political Economy* 129, no. 10: 2861–904.

Iacoviello, Matteo. 2005. "House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle." *American Economic Review* 95, no. 3: 739–64.

Joint Center for Housing Studies (JCHS). 2020. *The State of the Nation's Housing 2020*. Cambridge, Mass.: Harvard University. https://www.jchs.harvard.edu/state-nations-housing-2020.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2015. "Leveraged Bubbles." *Journal of Monetary Economics* 76:S1–S20.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2020. "The Effects of Quasi-Random Monetary Experiments." *Journal of Monetary Economics* 112:22–40.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018. "Monetary Policy According to HANK." *American Economic Review* 108, no. 3: 697–743.

Kekre, Rohan, and Moritz Lenel. 2020. "Monetary Policy, Redistribution, and Risk Premia." Working Paper 2020-02. Chicago: University of Chicago, Becker

Friedman Institute for Economics. https://bfi.uchicago.edu/wp-content/uploads/BFI_WP_202002.pdf.

Kent, Ana Hernandez, and Lowell Ricketts. 2021. "Wealth Gaps between White, Black and Hispanic Families in 2019." Blog post, January 5, Federal Reserve Bank of St. Louis On the Economy. https://www.stlouisfed.org/on-the-economy/2021/january/wealth-gaps-white-black-hispanic-families-2019.

Kuhn, Moritz, and José-Víctor Ríos-Rull. 2016. "2013 Update on the US Earnings, Income, and Wealth Distributional Facts: A View from Macroeconomics." *Federal Reserve Bank of Minneapolis Quarterly Review* 37, no. 1: 2–73.

Kuhn, Moritz, Moritz Schularick, and Ulrike I. Steins. 2020. "Income and Wealth Inequality in America, 1949–2016." *Journal of Political Economy* 128, no. 9: 3469–519.

Lenza, Michele, and Jirka Slacalek. 2021. "How Does Monetary Policy Affect Income and Wealth Inequality? Evidence from Quantitative Easing in the Euro Area." Discussion Paper 16079. London: Centre for Economic Policy Research.

Mäki-Fränti, Petri, Aino Silvo, Adam Gulan, and Juha Kilponen. 2022. "Monetary Policy and Inequality: The Finnish Case." Working Paper. https://helda.helsinki.fi/bof/bitstream/handle/123456789/18277/BoF_DP_2203.pdf?sequence=1.

Moll, Benjamin. 2020. Comment on "Sources of US Wealth Inequality: Past, Present, and Future." *NBER Macroeconomics Annual* 35:468–79.

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

Paiella, Monica. 2009. "The Stock Market, Housing and Consumer Spending: A Survey of the Evidence on Wealth Effects." *Journal of Economic Surveys* 23, no. 5: 947–73.

Paul, Pascal. 2020. "The Time-Varying Effect of Monetary Policy on Asset Prices." *Review of Economics and Statistics* 102, no. 4: 690–704.

Poterba, James M. 2000. "Stock Market Wealth and Consumption." *Journal of Economic Perspectives* 14, no. 2: 99–118.

Powell, Jerome. 2020. "New Economic Challenges and the Fed's Monetary Policy Review." Speech given at Navigating the Decade Ahead: Implications for Monetary Policy, Federal Reserve Bank of Kansas City, August 27. https://www.federalreserve.gov/newsevents/speech/powell20200827a.htm.

Ramey, Valerie A. 2016. "Macroeconomic Shocks and Their Propagation." In *Handbook of Macroeconomics, Volume 2*, edited by John B. Taylor and Harald Uhlig. Amsterdam: North-Holland.

Rigobon, Roberto, and Brian Sack. 2004. "The Impact of Monetary Policy on Asset Prices." *Journal of Monetary Economics* 51, no. 8: 1553–75.

Rodgers, William M. 2008. "African American and White Differences in the Impacts of Monetary Policy on the Duration of Unemployment." *American Economic Review* 98, no. 2: 382–86.

Romer, Christina D., and David H. Romer. 2004. "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review* 94, no. 4: 1055–84.

Rungcharoenkitkul, Phurichai, Claudio Borio, and Piti Disyatat. 2021. "Monetary Policy Hysteresis and the Financial Cycle." Working Paper. Basel: Bank for International Settlements. https://www.bis.org/publ/work817.htm.

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.

Thompson, Jeffrey P., and Gustavo Suarez. 2017. "Updating the Racial Wealth Gap." Working Paper 2015-76. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/econres/feds/updating-the-racial-wealth-gap.htm.

Wachtel, Paul, and Mario I. Blejer. 2020. "A Fresh Look at Central Bank Independence." *Cato Journal* 40, no. 1: 105–32.

Wolff, Edward N. 2016. "Household Wealth Trends in the United States, 1962 to 2013: What Happened Over the Great Recession?" *Russell Sage Foundation Journal of the Social Sciences* 2, no. 6: 24–43.

Wolff, Edward N. 2017. "Household Wealth Trends in the United States, 1962 to 2016: Has Middle Class Wealth Recovered?" Working Paper 24085. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24085.

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.

Zavodny, Madeline, and Tao Zha. 2000. "Monetary Policy and Racial Unemployment Rates." *Federal Reserve Bank of Atlanta Economic Review* 85, no. 4: 1–16.

Zonta, Michela. 2019. "Racial Disparities in Home Appreciation." Center for American Progress. https://www.americanprogress.org/issues/economy/reports/2019/07/15/469838/racial-disparities-home-appreciation/.

Comments and Discussion

COMMENT BY

WILLIAM A. DARITY JR. Bartscher, Kuhn, Schularick, and Wachtel partition the distributional effects of monetary policy into an income effect driven by employment and earnings effects and a wealth effect driven by portfolio and capital account effects. Focusing primarily on accommodative (or expansionary) monetary policy, they conclude that racial inequality in income will decrease while racial inequality in wealth will increase. Because Black households have considerably fewer financial assets than white households, the wealth gain is greater for white households—and their portfolio gain outweighs the relative income gain for Black households.

To conduct their analysis, the authors utilize the Romer-Romer shock series to assess the consequences of a comparatively incremental 1 percentage point (100 basis point) decrease in the rate of interest on Black-white economic disparities. The result of central bank policy in their simulation exercise, while not inconsequential, is not large. For example, they find that the net difference in consumption from capital gains for white households over Black households will total $470 over five years, or less than $8 a month.

I assume their motivation for using the shock series is due to their embrace of the new classical macroeconomics proposition that only unanticipated policy moves can have real effects on the economy's performance. Personally, I am skeptical about the proposition in a world where speculation rules financial markets. Announced changes in interest rates by the Federal Reserve immediately have non-sterilizing ripple effects across asset markets.

Moreover, the authors do not run their simulation exercise for restrictive (or contractionary) monetary policy; instead, they refer primarily to studies conducted on data from Scandinavian countries where the racial

Brookings Papers on Economic Activity, Spring 2022: 48–63

comparisons are not feasible. Therefore, it is not clear what their model implies for the distributional effects of Federal Reserve–engineered increases in interest rates. Is there simply a reversal of the effects of accommodative policy to a similar degree? We simply do not know based upon this study, although, as I will suggest below, the effects of accommodative and contractionary policies probably are not symmetrical.

Three measurement issues point toward limitations in this study. First, the authors rely upon the difference at the median households to calculate the magnitude and assess the policy effect on the Black-white wealth gap. This choice is conventional and generally seen as uncontroversial because the medians are immune to outlying values and, presumably, are more representative of the typical experience of most households in each group.

However, the wealth gap should be evaluated at the mean, rather than the median, leading to a Black-white net worth difference more than five times larger—and implying smaller repercussions for racial wealth inequality from *any* policy. The mean difference is preferable for estimation of the racial wealth gap because of the high degree of concentration of wealth in the United States.

Ninety-seven percent of white wealth is held by white households with a net worth above the median. Use of the median gap instead of the mean ignores an overwhelming share of wealth held by white households. Furthermore, the immense concentration of wealth among white households with a net worth above the white median is not due merely to the presence of a small number of white billionaires. One-quarter of white households have a net worth above $1 million; the same is true for only 4 percent of Black households (Darity, Addo, and Smith 2021). Plus, median net worth of members of the white working class consistently is two to three times as high as median net worth of members of the Black professional-managerial class (Addo and Darity 2021).

A focus on inequality in income or wealth does not necessarily provide a sufficient gauge of disparities in well-being. An important source of racial income inequality is racially uneven unemployment rates. The Black-white unemployment rate ratio is an index of the degree of discrimination in the economy (Wilson and Darity 2022), but it is not adequate as an index of relative deprivation.

Historically the ratio of Black to white unemployment rates consistently has been 2:1. The proportion is virtually a stylized fact. A rare exception occurred in April 2020, the first month after the onset of the pandemic in the United States. The Black to white unemployment rate ratio nearly reached parity at 1.1:1 (16.7 percent for Black workers and 14.2 percent

for white workers). Subsequently it reverted to the historical norm. Indeed, as of April 2022 the Black rate was 5.9 percent and white rate was 3.2 percent. Still, conditions were considerably worse for Black Americans in April 2020 due to the much higher unemployment rates, independent of the ratio.

There are issues with using the wealth ratio as a guide to relative well-being, but not for precisely the same reason. Black wealth is so low in comparison with white wealth that marginal increases in the Black-white ratio may have no substantive impact in the lived experiences of Black Americans.

Although the authors' conclusion about the impact of accommodative monetary policy on the degree of racial inequality is plausible, there are reasons to doubt that the effect of restrictive policy is symmetrical. After all, it is always possible for a central bank to conduct the latter policy, but it is not always possible to conduct the former. Monetary authorities always can press the brakes by raising interest rates higher, but they cannot always step on the accelerator by lowering interest rates.

Consider the case of Japan under conditions of very slow growth and extremely low interest rates (effectively negative real rates). It would not be possible for the central bank to reduce the nominal rate of interest by 100 basis points in that circumstance since the nominal rate already is less than one. The impact of monetary policy on overall distribution or intergroup distribution of resources is unlikely to be of any significance because conventional expansionary central bank interest management is empty. Certainly, while it would be feasible, it would not be desirable to *raise* interest rates in this situation.

The key to the absence of real effects from accommodative monetary policy is not whether the policy is anticipated or unanticipated. The key is the initial conditions under which the policy is introduced, and the authors do not consider the impact of interest rate management under various initial conditions.

Nor do they consider non-incremental changes in monetary policy, and distributional effects are of greater concern with non-incremental changes. In the early 1980s the Federal Reserve under Paul Volcker's leadership pursued unprecedentedly tight monetary policy. The federal funds rate soared from 9 percent to 19 percent as the Federal Reserve induced a major recession while suppressing inflation. In 1982–1983, the Black unemployment rate exceeded 20 percent while the white rate hit a ceiling just above 10 percent. Patently there were huge adverse effects on Black employment and income, far larger than the positive effects displayed in the authors' simulation (Sablik 2013).

Indeed, I contend that the impact on racial inequality is more unsettling when the Federal Reserve pursues tight monetary policy when inflation is present or imminent than when the Federal Reserve pursues loose interest rate policy to stimulate the economy. It is clear that a Volcker-style aggressive use of the Federal Reserve's traditional instruments for fighting inflation can produce a savage downturn, with harmful effects for most Americans, but disproportionately so for Black Americans.

Hence, I advocate a federal job guarantee—an assurance of access to public sector employment for all Americans at non-poverty wages with a benefits package similar to the one received by federal civil servants. De facto, it would make the Phillips curve vertical at a zero unemployment rate; no one would be looking for work who could not find it. The Federal Reserve could tackle inflation knowing the public at large is insulated from the effects of a recession.

In the absence of the job guarantee as a form of social insurance plan, we best beware the effects of non-accommodative monetary policy.

Moreover, from the standpoint of the Black community, inflation can have disproportionate adverse effects (Lalljee 2021). However, a recession induced by the Federal Reserve's anti-inflation measures also is likely to have disproportionate adverse effects on Black Americans, for example, the racial wage gap widens as Black compensation falls more than white compensation (Chattopadhyay and Bianchi 2020). If we want to unbridle the Federal Reserve's inflation-fighting capabilities, we must minimize the repercussions of an economic downturn. A federal job guarantee is essential toward that end.

REFERENCES FOR THE DARITY COMMENT

Addo, Fenaba, and William Darity Jr. 2021. "Disparate Recoveries: Wealth, Race, and the Working Class after the Great Recession." *Annals of the American Academy of Political and Social Science* 695, no. 1: 173–92.

Chattopadhyay, Shinjinee, and Emily C. Bianchi. 2020. "Does the Black/White Wage Gap Widen during Recessions?" *Work and Occupations* 48, no. 3: 247–84.

Darity, William, Jr., Fenaba Addo, and Imari Z. Smith. 2021. "A Subaltern Middle Class: The Case of the Missing 'Black Bourgeoisie' in America." *Contemporary Economic Policy* 39, no. 3: 494–502.

Lalljee, Jason. 2021. "People of Color and Rural Households Are Getting Hit the Hardest by the Hot Inflation of 2021." *Business Insider*, November 24. https://www.businessinsider.com/inflation-hurts-black-latino-hispanic-people-of-color-rural-households-2021-11.

Sablik, Tim. 2013. "Recession of 1981–82." Federal Reserve History. https://www.federalreservehistory.org/essays/recession-of-1981-82.

Wilson, Valerie, and William Darity Jr. 2022. *Understanding Black-White Disparities in Labor Market Outcomes Requires Models that Account for Persistent Discrimination and Unequal Bargaining Power*. Washington: Economic Policy Institute. https://www.epi.org/unequalpower/publications/understanding-black-white-disparities-in-labor-market-outcomes/.

COMMENT BY

BENJAMIN MOLL Bartscher, Kuhn, Schularick, and Wachtel approach a controversial question with high-quality empirical evidence. The result is a very valuable contribution to the literature on both monetary policy and racial inequality. To make this contribution, the authors combine household balance sheet data for Black and white households since the 1950s from an impressive data compilation effort by Kuhn, Schularick, and Steins (2020) with time series estimates of the response of asset prices and unemployment rates to monetary policy shocks.

The authors' main finding is that interest rate cuts have opposite effects on racial income and wealth inequality. On the one hand, they decrease the racial unemployment gap and therefore the percentage gap between Black and white earnings. On the other hand, they increase asset prices and therefore the racial wealth gap. The key to the latter result is large and very persistent estimated asset price increases in response to interest rate cuts (the instrumental variable local projections results in their figure 8) and that white households hold portfolios that are more concentrated in assets with rising prices such as equities. The authors conclude that "monetary policymakers face a trade-off: monetary accommodation widens racial wealth inequality as it reduces income inequality."

In parts of the paper, the authors compare the size of these earnings and wealth effects (see, e.g., their figure 1) and advance a more provocative version of this conclusion, namely, that "the reduction in the earnings gap pales in comparison to the effects on the wealth gap" and that "our analysis therefore does not bode well for the suggestion . . . that a more accommodative monetary policy helps alleviate racial inequalities."

Much of my comment will focus on the question whether and to what extent unrealized capital gains generated by falling interest rates are comparable to earnings changes. I will first draw on discussions of this and related issues in the last one hundred years of economic thought. I will then comment on a strategy the authors use for making this comparison, namely, to translate capital gains into consumption effects. Finally, I will draw on some of my own work that is relevant to the question at hand (Fagereng and others 2022).

APPLES VERSUS ORANGES? At various points in the paper, Bartscher, Kuhn, Schularick, and Wachtel compare the earnings gains from accommodative monetary policy with the corresponding capital gains. This is, of course, an easy comparison to make: after all, both quantities are in dollars. However, a naïve comparison like that in the authors' figure 1 risks overlooking an important fact: the capital gains calculated by the authors are *unrealized* capital gains and do not automatically yield higher cash flows, disposable income, or consumption. This raises the question whether this comparison in fact amounts to a comparison of apples and oranges? Put differently: Are unrealized capital gains income? In particular, those generated by a decline in interest rates?

Unrealized capital gains in the history of economic thought. The question whether unrealized capital gains are income has a long tradition in economics, going back to work by Haig (1921) and Simons (1938). In their work the answer is yes: their proposed income definition—which is now known as "Haig-Simons income"—includes unrealized capital gains. A frequent next step in this line of argument is that capital gains should be taxed on accrual rather than realization.

However, this view soon received pushback, for example, in Nicholas Kaldor's classic book *An Expenditure Tax* (1955).[1] One excerpt is worth citing: "We may now turn to the other type of capital appreciation which reflects a fall in interest rates rather than the expectation of higher earning power. This in a sense is in an intermediate category . . . since the rise in capital values in this case [comes] without a corresponding increase in the flow of real income accruing from that wealth" (44). This has the following implication: "For in so far as a capital gain is realized and spent . . . the benefit derived from the gain is equivalent to that of any other casual profit. If however it is not so realized, there is clearly only a smaller benefit" (ibid.).

Kaldor's message is clear: if earnings are apples, unrealized capital gains due to declining interest rates are oranges. This difficulty, along with some other ones, ends up critically influencing Kaldor's thinking about taxation: he concludes that it is exceedingly difficult to define a notion of income that would form a good tax base. Kaldor's solution: an expenditure tax—hence the title of his book.

Since these early contributions, a sizable and growing theoretical and quantitative literature in macroeconomics and household finance has examined the effect of asset price changes on wealth and welfare inequality.

1. See also Paish (1940), who does not reference Haig (1921) and Simons (1938) but makes a similar point.

See, for example, Whalley (1979), Gomez (2019), Gomez and Gouin-Bonenfant (2020), Catherine, Miller, and Sarin (2020), İmrohoroğlu and Zhao (2022), Moll (2020), Greenwald and others (2021), Cioffi (2021), Fagereng and others (2019), and Fagereng and others (2022). Several contributions pursue lines of argument similar to Kaldor's, namely, that unrealized capital gains, in particular those due to falling interest rates, should be treated differently from income gains. I especially recommend Whalley (1979), who provides a beautifully clear graphical analysis of a two-period model similar to that in Moll (2020), as well as the nontechnical expositions by Cochrane (2020) and Krugman (2021).

The authors' solution: from capital gains to consumption. Bartscher, Kuhn, Schularick, and Wachtel are, of course, aware of the difficulty involved in comparing earnings gains with unrealized capital gains. They therefore propose to look at the consumption effects of capital gains rather than the capital gains themselves. The idea is simple: if households realize their capital gains (i.e., if they sell the asset whose price appreciates) and consume the proceeds, this gets around the apples versus oranges problem.

Unfortunately, the authors' SCF+ data do not feature information on households' consumption. They therefore use an estimate for the marginal propensity to consume (MPC) out of capital gains from the literature, namely, a cross-regional estimate of 3.2 percent from Chodorow-Reich, Nenov, and Simsek (2021), and apply this estimate to the wealth gains of both Black and white households. Thus, a wealth gain for white households of \$18,900 yields a consumption gain of $.032 \times \$18{,}900 = \$605$, and a wealth gain for Black households of \$3,300 yields a consumption gain of $.032 \times \$3{,}300 = \$105$, so that the relative consumption gain for white households is $\$605 - \$105 = \$500$, which is considerably larger than the accumulated relative earnings effect for Black households of \$134. A practical difficulty is that the literature does not feature separate estimates of such MPCs out of capital gains by race. The authors address this shortcoming with a simple back-of-the-envelope calculation: they calculate how large differences in Black and white MPCs would have to be in order to offset the relative income effect. They find that "Black MPCs would have to be roughly three times larger than white MPCs for the relative consumption effect from capital gains to be as large as the relative earnings effect" and argue that such large differences in Black versus white MPCs are implausible.

This is a very useful and convincing line of reasoning. Given the data constraints faced by the authors, in particular the absence of consumption information, it is probably also the best they can do. In my view, however, it is still not fully satisfactory. My main hesitation is that the estimated

consumption effect is the outcome of multiplying three numbers estimated using completely separate data sets: household balance sheet positions from the SCF+ are multiplied with time series estimates of asset price responses to monetary policy shocks to get the wealth gains; these are then further multiplied with an MPC estimate from the cross-regional analysis in Chodorow-Reich, Nenov, and Simsek (2021) to get consumption effects. It is therefore unclear to what extent interest rate cuts actually translate into higher asset prices and higher consumption for the households in the SCF+.

Alternative solution: equivalent variation of asset price changes. It is therefore worth asking: Are there any alternative ways to "translate the oranges into apples"? In recent work, Fagereng and others (2022) show that the answer is yes: one can translate asset price changes due to interest rate cuts into a money metric welfare measure that is comparable to income gains. More precisely, we provide a sufficient statistics formula for the equivalent variation of asset price changes. In the context of rising asset prices due to interest rate cuts, the formula answers the following question: What is the equivalent transfer the individual would have to receive to experience the same welfare change as from the asset price increase?[2] The formula takes the following form:

$$Welfare\ Gain_i = \sum_{t=0}^{T} R^{-t} \left(Sales_{it} \times Price\ Deviation_t \right)$$

where i denotes the individual, T is the length of the sample period, $R > 1$ is a discount rate, *Salesit* are the net sales of the asset by the individual in year t, and $Price\ Deviation_t$ is the deviation of the asset price due to interest rate cuts. Under some additional assumptions, this price deviation can be computed as the percentage change over time in the asset's price–dividend ratio:

$$Price\ Deviation_t = \Delta\% \left(\frac{Price}{Dividend} \right)_t .$$

Intuitively, an interest rate cut is an example of a discount rate shock, as in Campbell and Shiller (1988). It thus leads to an increase in the asset's

2. Our sufficient statistics formula is a first-order approximation to the welfare gain and so the equivalent variation also equals the compensating variation, that is, the net revenues of a planner who must compensate the individual for the asset price deviation, bringing the individual back to their welfare in the baseline scenario. The formula shown here is for the case of one asset; Fagereng and others (2022) extend it to multiple assets in a straightforward manner.

valuations as measured by its price–dividend ratio. The formula follows from an application of the envelope theorem and thus holds to first order.[3]

The formula for welfare gains generates two main insights. First, what matters are asset transactions, not asset holdings. Intuitively, higher valuations are good news for prospective sellers (those with $Sales_{it} > 0$) and bad news for prospective buyers (those with $Sales_{it} < 0$). A particularly interesting case is an individual who owns assets but does not plan to buy or sell ($Sales_{it} = 0$). For such an individual, rising asset prices are merely "paper gains," with no corresponding welfare implications. Second, asset price changes are purely redistributive. When asset prices rise, there is a redistribution of welfare from sellers to buyers. But since for every seller there is a buyer, summing the formula across all parties and counterparties of financial transactions in the economy implies that the welfare gains aggregate to zero.[4]

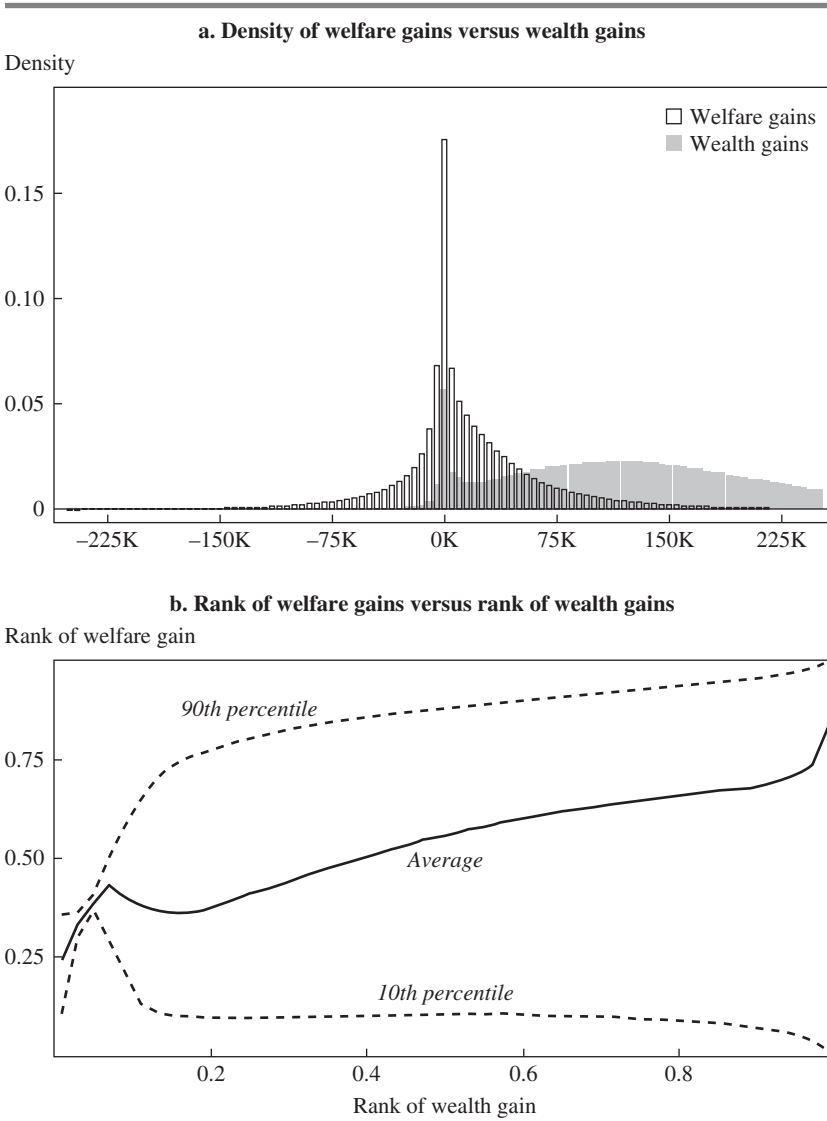
In Fagereng and others (2022) we operationalize this approach using Norwegian administrative panel data on asset transactions for the time period 1994 to 2015 so as to identify the winners and losers of historical asset price changes over this time period (all asset price changes, not just those due to monetary policy). The result is the histogram labeled "Welfare gains" in figure 1, panel a. As expected, the figure shows substantial dispersion, that is, there are some large winners and some large losers, reflecting large sales and purchases; at the same time, the welfare gains and losses are centered around zero, reflecting the purely redistributive role of asset price changes.

An important question is: How do these welfare gains compare to wealth gains from rising asset prices, that is, the unrealized capital gains emphasized by Bartscher, Kuhn, Schularick, and Wachtel? The second histogram in figure 1, panel a, labeled "Wealth gains," plots exactly this quantity. The main observation is that while welfare gains are centered around zero, wealth gains are centered at a large positive value. This reflects the fact that wealth gains accrue to all asset holders while welfare gains only accrue to asset sellers.

3. The formula omits an effect that may be important in practice: that rising asset prices loosen collateral constraints, thereby allowing for more borrowing and consumption. The formula can be extended to take this effect into account.

4. However, since there are financial transactions between sectors of the economy (i.e., between households, the government, and foreigners), we can have a case in which the household sector as a whole benefits, but necessarily at the expense of another sector. In our empirical implementation in Fagereng and others (2022), we find that the welfare gains of the Norwegian household sector approximately aggregate to zero.

Figure 1. Comparing Welfare Gains (Equivalent Variation) and Wealth Gains



Source: Fagereng and others (2022); reproduced with permission.

While this exercise shows that welfare and wealth gains have different densities, it is silent on the correlation between the two variables. To focus on this question, figure 1, panel b, plots the average rank of welfare gains versus the rank of wealth gains. If welfare gains are perfectly correlated with wealth gain, the result would be a 45 degree line from zero to 1. Conversely, if welfare gains are unrelated to wealth gain, the result should be a horizontal line at 0.5. Reality is somewhere in-between: empirically, some individuals with large asset positions buy and hence lose in welfare terms; conversely, others with small positions sell and hence win. This finding also shows up in the wide bands for the 10th and 90th percentile welfare gains: within any given wealth gain rank, some individuals experience a very low welfare gain and others experience a very high one.

CONCLUSION The implications of these considerations for the work of Bartscher, Kuhn, Schularick, and Wachtel are clear: care is needed when comparing earnings gains and unrealized capital gains. In my view, a fully satisfactory comparison would require a data set with information on three variables: either all of household income, wealth, and *consumption*; or all of household income, wealth, and *asset transactions*. Unfortunately, neither is available in the United States.[5]

Given the difficulty of comparing earnings gains and unrealized capital gains as well as the data limitations in the United States, the authors' provocative conclusion that accommodative monetary policy hurts overall racial inequality should be taken with a grain of salt. At the same time, the less provocative part of their conclusion stands: monetary policy seems to face a trade-off with respect to racial inequality in that interest rate cuts widen racial wealth inequality as they reduce earnings inequality. This is an important finding and constitutes a very valuable contribution in itself.

REFERENCES FOR THE MOLL COMMENT

Campbell, John Y., and Robert J. Shiller. 1988. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies* 1, no. 3: 195–228.

Catherine, Sylvain, Max Miller, and Natasha Sarin. 2020. "Social Security and Trends in Wealth Inequality." Working Paper. Philadelphia: Jacobs Levy Equity Management Center for Quantitative Financial Research. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3546668.

5. The Panel Study of Income Dynamics (PSID) does feature information on all household income, wealth, and consumption. However, it has a number of downsides, including that it surveys households only every two years, potential issues with the reliability of the consumption data, and that it does not accurately capture the very top of the income and wealth distributions.

Chodorow-Reich, Gabriel, Plamen T. Nenov, and Alp Simsek. 2021. "Stock Market Wealth and the Real Economy: A Local Labor Market Approach." *American Economic Review* 111, no. 5: 1613–57.

Cioffi, Riccardo A. 2021. "Heterogeneous Risk Exposure and the Dynamics of Wealth Inequality." Working Paper. https://www.rcioffi.com/files/jmp/cioffi_jmp2021_princeton.pdf.

Cochrane, John. 2020. "Wealth and Taxes, Part II." Blog post, January 3, The Grumpy Economist. https://johnhcochrane.blogspot.com/2020/01/wealth-and-taxes-part-ii.html.

Fagereng, Andreas, Matthieu Gomez, Émilien Gouin-Bonenfant, Martin Holm, Benjamin Moll, and Gisle Natvik. 2022. "Asset-Price Redistribution." Working Paper. https://www.matthieugomez.com/files/apr.pdf.

Fagereng, Andreas, Martin Blomhoff Holm, Benjamin Moll, and Gisle Natvik. 2019. "Saving Behavior across the Wealth Distribution: The Importance of Capital Gains." Working Paper 26588. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26588.

Gomez, Matthieu. 2019. "Asset Prices and Wealth Inequality." Working Paper. https://www.matthieugomez.com/files/wealthinequality.pdf.

Gomez, Matthieu, and Émilien Gouin-Bonenfant. 2020. "A Q-Theory of Inequality." Working Paper. https://www.matthieugomez.com/files/qtheory.pdf.

Greenwald, Daniel L., Matteo Leombroni, Hanno Lustig, and Stijn Van Nieuwerburgh. 2021. "Financial and Total Wealth Inequality with Declining Interest Rates." Working Paper 28613. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28613.

Haig, Robert Murray. 1921. "The Concept of Income: Economic and Legal Aspects." In *The Federal Income Tax*, edited by Robert Murray Haig. New York: Columbia University Press.

İmrohoroğlu, Ayşe, and Kai Zhao. 2022. "Rising Wealth Inequality: Intergenerational Links, Entrepreneurship, and the Decline in Interest Rate." *Journal of Monetary Policy* 127:86–104.

Kaldor, Nicholas. 1955. *An Expenditure Tax*. London: Routledge.

Krugman, Paul. 2021. "Pride and Prejudice and Asset Prices." *New York Times*, October 5. https://www.nytimes.com/2021/10/05/opinion/low-interest-rates-monetary-policy.html.

Kuhn, Moritz, Moritz Schularick, and Ulrike I. Steins. 2020. "Income and Wealth Inequality in America, 1949–2016." *Journal of Political Economy* 128, no. 9: 3469–519.

Moll, Benjamin. 2020. Comment on "Sources of US Wealth Inequality: Past, Present, and Future." *NBER Macroeconomics Annual* 35:468–79.

Paish, F. W. 1940. "Capital Value and Income." *Economica* 7, no. 28: 416–18.

Simons, Henry Calvert. 1938. *Personal Income Taxation: The Definition of Income as a Problem of Fiscal Policy*. Chicago: University of Chicago Press.

Whalley, John. 1979. "Capital Gains Taxation and Interest Rate Changes: An Extension of Paish's Argument." *National Tax Journal* 32, no. 1: 87–91.

GENERAL DISCUSSION Justin Wolfers pointed out that by relying on the Romer-Romer shocks for identification, the paper ends up focusing exclusively on when the Federal Reserve deviates from a Taylor-type rule—that is, when the Federal Reserve does something unusual. But, he continued, the paper is silent on the implications of normal systematic policy for racial gaps. Wolfers argued that we know that unexpected shocks move financial markets a lot, and we would expect the unsystematic part of monetary policy to have a large effect on wealth, while the systematic part would not. But if we think about income, we might expect the same effect on the racial gap whether unanticipated or not. He concluded by suggesting that one might get very different implications from regular monetary policy for the racial gap than when analyzing only the shocks.

Henry Aaron pondered whether the five-year horizon in the paper is long enough. He suggested that we would expect a drop in interest rates to increase the value of assets for a given income flow, given the change in discounting of the future, which would be realized relatively quickly, and if we change the underlying labor market conditions, there may be some initial effect on employment—especially at the bottom end of the income distribution. But then, over time, changes such as opportunities for promotion and investment in human capital by employees as well as employers would result in a continuing effect on earnings but, after an initial period, little or no additional effect on asset values.

Frederic Mishkin raised the issue of whether Federal Reserve policy should take into account effects on racial inequality and argued that the answer is no. He suggested an approach where we may want to think about nonmonetary policies that can counteract potentially harmful outcomes of Federal Reserve policy instead, where in addition to the important issue of racial inequality there are more general income and distributional effects which the government can address. Mishkin noted that there are many good reasons the Federal Reserve should be less involved in these very political issues and that the Federal Reserve only has one instrument and already has two objectives it has to focus on: employment and inflation.

Responding to Mishkin's comment, Wolfers pointed out that the Federal Reserve already expresses concern about the effect of its policies on savers and on borrowers—Wall Street gets listened to—and that it seems hard to rationalize that the Federal Reserve should be responsive to these groups, but not when it comes to racial disparities.

Mishkin said that the Federal Reserve should *not* be responsive to the special interests of Wall Street either but rather needs its independence to deal with unemployment and inflation.

Paul Wachtel commented that he agrees that racial inequality should not be another goal of monetary policy, but that the Federal Reserve always is—and should be—mindful of what else is going on in the economy and the implications of its policy. He said that many would agree if we were talking about financial stability but argued that racial inequality should also be included in its deliberations. He added that being mindful of these things does not mean that accommodative policy is or is not appropriate at any given time. Wachtel concluded by suggesting that monetary policy is more complex than economists may have thought thirty years ago.

William Darity wondered if what is really meant is that the Federal Reserve should be independent not of politics but of partisanship? He argued that it would be virtually impossible for monetary policy to not be political and that which factors are taken into consideration when the Federal Reserve makes decisions and what the implications are is always a question that comes into play. Darity suggested that adding an additional consideration would not cause a fundamental problem. He agreed that certain types of social issues cannot be addressed effectively with monetary policy and that the same argument could be made for racial wealth inequality: that the only way in which the racial wealth gap in the United States can truly be addressed is through reparations.

Commenting on Aaron's remarks, Betsey Stevenson agreed that we need also to consider consumption in the future. High stocks of wealth and lower interest rates can provide a lot of opportunities in the future. In the labor market, we have to think about how extended spells of unemployment have effects that last for a very long time. On Mishkin's remarks, Stevenson pointed to the trade-offs that come with running accommodative monetary policy and noted that the paper considers the risks of high inflation, which will hurt savers but which at the same time allows for the accumulation of a lot of wealth. She pondered whether this offsetting effect should be taken into account when considering the risks of running accommodative policy.

Donald Kohn said that the paper confirmed his intuition that accommodative monetary policy reduces income inequality while raising wealth inequality. He agreed with Mishkin that monetary policy is not the right tool to address income and wealth inequality. Even if accommodative policy would increase racial income gaps, it would lower unemployment—and why would you want to deprive any household, Black or white, of employment? He pondered what the broad-based and inclusive goal of the Federal Reserve means relative to the goal of maximum employment consistent with stable prices and said he believed there is no stable trade-off between the racial gaps and the employment and inflation objectives.

Stefania Albanesi made the point that cross-country evidence tells us that high inflation causes unfavorable redistribution for low-income households, through their balance sheets but also because they experience higher inflation.[1] Albanesi noted that inflation did not come up in the paper but wondered whether a high-inflation environment would have had a different impact on redistribution than the low-inflation environment and the policy that comes with it, which is what we have been seeing for the past many decades.

Ben Bernanke commented in the virtual conferencing chat that a naïve reader of the paper may conclude that in the interest of racial equality the Federal Reserve should never ease monetary policy, even in a deep recession. He asked whether this was indeed the correct inference from the findings in the paper.

Moritz Schularick responded that the paper gains may not be as inconsequential as one might think. In a world with borrowing constraints, resulting collateral constraints may have a permanent effect on wealth inequality by reducing or increasing opportunities, including opportunities for starting a business or making human capital investments. Schularick argued that realized capital gains can be connected to systematic distributional consequences and noted that previous research by Glover and others shows that during the Great Recession, for welfare reasons, younger households systematically preferred asset prices dropping more than wages.[2] He suggested that a parallel systematic argument can be made with respect to racial inequalities: some groups may profit, and others may not. Also, if you have inherited inequalities in something like homeownership as a result of discrimination in the past, sellers and buyers will be two distinct groups—these systematic differences matter and are worth thinking about. Ultimately, he agreed with Bernanke's bottom line: there are important tradeoffs that merit more consideration.

Pushing back on the point made by Benjamin Moll in his comment on the paper—that the marginal propensity to consume (MPC) across Black and white households would significantly affect the results—Schularick argued that one would have to make pretty extreme assumptions of the difference in MPC for the results of the paper to be reversed.

1. See Stefania Albanesi, "Inflation and Inequality," *Journal of Monetary Economics* 54, no. 4 (2007): 1088–114.

2. Andrew Glover, Jonathan Heathcote, Dirk Krueger, and José-Víctor Ríos-Rull, "Intergenerational Redistribution in the Great Recession," *Journal of Political Economy* 128, no. 10 (2020): 3730–78.

Moll made a plea for better data collection—comprehensive data on income, wealth, and consumption expenditure are needed and do not currently exist in the United States, hampering efforts to progress our understanding on these matters.

Austan Goolsbee commented in the chat that it can be interesting to think of the geographic incidence of monetary policy in the same way as for sectors or racial groups. He suggested that the question of whether the Federal Reserve should pay attention to a hard-hit group specifically or just the aggregate becomes similar to the issue of optimal currency areas and whether the central bank should respond to shocks hitting one geographic area differently than others.

CLAUDIA GOLDIN
Harvard University

Understanding the Economic Impact of COVID-19 on Women

ABSTRACT Compared with previous recessions, the recession induced by COVID-19 had a greater impact on women's employment and labor force participation relative to men. But the big divide was less between men and women than it was between the more and the less educated. Contrary to many accounts, women did not exit the labor force in large numbers, and they did not greatly decrease their hours of work. The aggregate female labor force participation rate did not plummet. That said, the ability to balance caregiving and work differed greatly by education, occupation, and race. The more educated could work from home. Those who began the period employed in various in-person service occupations and establishments experienced large reductions in employment. Black women experienced a more negative impact beyond other factors considered, and the health impact of COVID-19 is a probable reason. The estimation of the pandemic's impact depends on the counterfactual used. The real story of women during the pandemic is that employed women who were educating their children and working adult daughters who were caring for their parents were stressed because they were in the labor force, not because they left.

The economic downturn that resulted from our self-induced COVID-19 coma has had economic effects different from those of any other recession or national crisis in US history. This time really has been different. Never before have we needed to shut down the economy to get it running again.[1]

1. The Great Influenza epidemic of 1918 was not accompanied or followed by an economic recession of any magnitude, possibly because of World War I or because the economy was not placed in as extreme a coma as ours has been. In addition, the virus may have rapidly mutated to a less virulent form, although not until after 1919.

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

Brookings Papers on Economic Activity, Spring 2022: 65–110

Never before have those working on the front lines been asked to bring danger back into their homes, not even in time of war. Never before inpeacetime have the caring sectors of education and health been as intertwined with the economic sectors of production and services. And never before has a recession had a greater impact on women than on men in a host of ways.

It seems safe to say that no one was untouched by the pandemic. But much of the deepest economic impact and personal pain was experienced by women. Many were caregivers for their own children and parents; some worked as aides for other people. Their jobs put them on the front lines in hospitals, nursing homes, and grocery stores. Others worked in restaurants, hospitality, retail, and personal service and saw their workplaces shuttered.

These women are of all types. But those who experienced the greatest impact were the mothers of school-age and younger children, Black and Hispanic women, single moms, and adult daughters who cared for parents. They may bear the marks of the pandemic for some time.

Women today are almost half of the total US labor force—47 percent just prior to March 2020, according to the Current Population Survey (CPS) data.[2] It is because of their great importance to the labor force that issues of caring and K-12 education took on greater significance during the pandemic and were seen as a means of restarting the economy. For the same reasons, uncertainty about the economic recovery was driven by the possibility that schools and day care centers would not fully open in fall 2021 or that parents would be fearful of sending their children to them. At the same time, it is because the vast majority of women 25 to 54 years old are in the labor force—76 percent in 2019—and half of them have children younger than age 18—that the care and education of children have been paramount to them.

The only time previously in US history that childcare was viewed as essential to the nation's economic health was during World War II. In 1940, only 18 percent of married women were in the labor force (Goldin 1991), and the overwhelming majority of Americans (both men and women born before 1930) agreed with the statement: "A preschool child would likely suffer if the mother worked for pay" (Goldin 2021).[3] But in 1942, firms

2. The fraction of the labor force by hours worked that is female is less than 47 percent since working women of all ages report working for pay 10 percent fewer hours than do working men (similarly for women age 25–54), using data for 2019 and "hours usually worked per week at all jobs," truncating at 84 hours.

3. From General Social Survey data for those born before 1930, although the question was asked in the 1970s and 1980s when the individuals were considerably older and they may have tempered their agreement with the statement.

across the nation encountered obstacles fulfilling contracts for war production. Firms entreated the federal government for the means to increase the labor supply with an untapped reserve—mothers.

The federal government responded by redeploying funds from the 1940 Lanham Act to set up thousands of nursery schools for children 2 to 5 years old and provide extended school programs for older children.[4] The policy appears to have done the trick. By 1944, 55 percent of servicemen's wives age 25–44 were in the labor force.[5]

Even in ordinary times, the care of children uses far more hours of mothers' time than of (custodial) fathers, and that is true even if both are college graduates and are fully employed. Thus, even though school and day care closings during the pandemic have had disproportionate impacts on most parents, the absolute time demands on mothers were extraordinary. By the estimates I will later offer, childcare (including education) time for children younger than age 18 increased from 8.7 hours per week before the pandemic, to 17.3 hours early in the pandemic, to around 22.4 hours by fall 2020 for college-graduate women (who were full-time workers with elementary school–age children in different sex, two-parent households). Childcare hours of custodial fathers also greatly increased early in the pandemic. But the increased hours of childcare of working women created an exceedingly heavy load (even weightier when ordinary housework is included).

The pandemic resulted in considerable burdens and stresses from its dual impact on people's health and jobs. The closing of schools and day care facilities, the furloughing of nannies and house cleaners, and the reduction of home health care workers intensified time demands on mothers and other women. Even though work from home was safer, it was filled with interruptions, prompting some to cleverly dub WFH as "work from hell."

It is no wonder that in the first year of the pandemic, especially in its first six months, news media and policy reports were jam-packed with alarming

4. On the Lanham Act, see Derrington, Huang, and Ferrie (2021). In a similar manner, many states used CARES Act funds to set up childcare facilities for the children of hospital workers, EMTs, first responders, and other workers deemed essential during the early months of the pandemic. These programs also served to buoy the hard-hit childcare sector.

5. According to official estimates, about 26 percent of all women (age 15 and older) were in the labor force in 1940. But by 1944—from estimates implicit in Goldin (1991)—the female labor force rapidly increased to 40 percent among those age 18 and older. It subsequently fell to 30 percent by 1947, about as rapidly as it had risen. It subsequently began its secular rise, reaching 34.5 percent by 1951. There has been no other time in recorded US history of so rapid a change in female labor force participation. We do not yet know whether the Lanham Act nursery schools had an impact on later female labor supply. On the role of mobilization rates on women's labor supply, see Goldin and Olivetti (2013).

headlines: "Pandemic Could Scar a Generation of Working Mothers" and "Pandemic Will 'Take Our Women 10 Years Back' in the Workplace."[6] A McKinsey-LeanIn Report that surveyed more than 40,000 employees from forty-seven companies issued a dire set of predictions, concluding that "1 in 4 women are contemplating . . . downshifting their careers or leaving the workforce" and "one in three mothers may be forced to scale back or opt out" (Thomas and others 2020, 6, 21). These prophecies became part of a media echo chamber, repeated again and again as if they had actually occurred.[7]

Even in March 2021, as vaccines were just entering arms of the under-65 crowd in most states, the news media continued to emphasize the reduction in women's employment and a future in which these reductions would be made permanent: "Pushed Out by Pandemic, Women Struggle to Regain Footing in U.S. Job Market" and "In One Year, Coronavirus Pandemic Has Wreaked Havoc on Working Women."[8]

Some even offered the disturbing possibility that female labor force participation has already dropped to 57 percent—levels not seen since 1988 (Forman 2022). One headline stated: "Women's Labor Force Participation Rate Hit a 33-Year Low in January," even though there was almost no sustained growth in that rate during those three decades (Connley 2021).

6. Patricia Cohen and Tiffany Hsu, *New York Times*, June 30, 2020, https://www.nytimes.com/2020/06/03/business/economy/coronavirus-working-women.html; Amanda Taub, *New York Times,* September 26, 2020, https://www.nytimes.com/2020/09/26/world/covid-women-childcare-equality.html.

7. The comments from the report that one in four women or one in three mothers might be leaving the workforce or cutting back were repeated in numerous news reports (for example, Kathryn Vasel, "The Pandemic Could Push an Alarming Number of Women Out of the Workforce," CNN Business, September 30, 2020, https://www.cnn.com/2020/09/30/success/women-senior-level-pandemic/index.html; Catherine Thorbecke, "1 in 4 Women Considering Leaving Workforce or Downshifting Careers because of COVID-19, Report Warns," ABC News, September 29, 2020, https://abcnews.go.com/Business/women-leaving-workforce-downshifting-careers-covid-19-report/story?id=73310740; Courtney Connley, "1 in 4 Women Are Considering Downshifting Their Careers or Leaving the Workforce due to the Coronavirus," CNBC, September 30, 2020, https://www.cnbc.com/2020/09/30/1-in-4-women-are-thinking-about-altering-their-careers-due-to-covid-19.html). Few, if any, also cited data in the McKinsey-LeanIn report showing that one in five fathers were also considering reducing hours or switching to a less demanding job (Thomas and others 2020, 21).

8. Jonnelle Marte and Aleksandra Michalska, *Reuters*, March 5, 2021, https://www.reuters.com/article/us-health-coronavirus-women-jobs/pushed-out-by-pandemic-women-struggle-to-regain-footing-in-u-s-job-market-idUSKBN2AW19Y; Tim Smart, *US News,* March 8, 2021, https://www.usnews.com/news/economy/articles/2021-03-08/in-one-year-coronavirus-pandemic-has-wreaked-havoc-on-working-women.

Versions of this headline were replayed across media outlets for months.[9] The enormous impact by race was also noted: "Taken together, the coronavirus proved to be a double whammy for Black women, robbing them of their jobs as well as threatening their health" (Smart 2021, par. 12). There was, unfortunately, considerable truth to that.

In the spring of 2021, we took off our masks and breathed in deeply, thinking we were on a straight road to economic recovery and health. We are less certain of that now, even as schools, day care centers, businesses, and offices have reopened. As the nation's labor force is slowly shifting to a new normal, it is prudent to look back and assess damage while exploring the potential for positive change.

In the year that followed, many researchers have examined the facts behind these headlines.[10] Although there is some variation in the findings, a consensus has developed around the economic impact of the pandemic on women and the veracity of these stories. I will summarize the main findings, explain some differences, and add a few additional considerations. My intention is to clarify the impact of the pandemic on women rather than evaluate the differential impact of the pandemic on women relative to men or relative to other recessions, a task capably accomplished by others.[11]

The consensus that has emerged regarding the actual economic impact of the pandemic on women generally includes the following five points, which are developed more fully below.

First, female labor force participation greatly increased in the half year preceding the pandemic, making judgments about changes after the pandemic began dependent on the starting point and the assumed counterfactual. The claim that the female labor force participation rate was rolled back to levels not seen for more than thirty years does not consider the fact

9. "This brings the total number of women who have left the labor force since February 2020 to more than 2.3 million, and it puts women's labor force participation rate at 57%, the lowest it's been since 1988" (Connley 2021, par. 2). "Now, 56 percent of American women are working for pay, the lowest level since 1986" (Miller 2021, par. 2).

10. The literature on the gendered features of the pandemic recession is now extensive. Among the pioneers in the scholarly literature who wrote on the subject as early as March 2020, Alon and others (2020a) were probably the first. That piece was followed by Alon and others (2020b), and then by a revealing work, Alon and others (2021), using comparable data for six countries (United States, Canada, Germany, Netherlands, Spain, and United Kingdom), exploring the impact of COVID-19 across nations with different social insurance systems. Heggeness (2020) was among the first to recognize the impact of leave-taking on employment.

11. See Albanesi (2021) and Albanesi and Kim (2021).

that female labor force participation had been flat for some time, and male labor force participation actually decreased. The female labor force participation rate, for those age 25–54 was about 75 percent in the early 1990s and has not been much different in more recent years. In fact, the rate was 75.6 percent before the pandemic in November 2018 and was also 75.6 percent in November 2021 (the last available month at the time of this writing).[12]

Second, compared with previous recessions, the one induced by COVID-19 affected women's employment and labor force participation somewhat more relative to men's and thus deserves the moniker "shecession." But the big divide is less between men and women and more between the more educated and the less educated. Although educational differences have been present in other recessions, the ability of the educated to work remotely and more safely would be expected to have exacerbated educational differentials relative to those in typical recessions.

Third, childcare time across all families with school-age and younger children probably doubled around the start of the pandemic. Childcare time for custodial fathers probably more than doubled for the first several months after March 2020, in part because their hours began at a lower level than those of custodial mothers. Childcare time for mothers probably increased further as some workplaces reopened in fall 2020 and custodial fathers reduced their childcare hours, yet schools did not remain open everywhere.

Fourth, labor market outcomes for women with young children were more affected than for others, but the pandemic had a great impact on all women. The employment and labor force participation of mothers with school-age and younger children varied by the mother's level of education and the year and season during the pandemic. An important finding is that employed mothers, by and large, did not leave the labor force despite their greatly increased time demands due to school and day care closings, and those who remained employed did not downshift as much as has been thought. But caregivers of children, the elderly, and the sick were burdened in a multitude of ways that became part of the media's headline stories.

Fifth, occupation and industry mattered considerably to women's employment. Prior to the pandemic, and relative to men, women were disproportionately employed in establishments—such as restaurants, retail, beauty salons, child day care services, and home health care services—that were shuttered or had restricted service by state mandates at the start of the pandemic. Even after they could open, these businesses had reduced demand,

12. Bureau of Labor Statistics, "Series Reports," https://data.bls.gov/cgi-bin/srgate, LNS11300062.

and many had permanently closed their doors. Race and ethnicity mattered to women's employment and labor force status independent of the age of their youngest child, occupation, and education.[13] Why that is the case may concern social disparities in COVID-19 health outcomes and the greater exposure of their jobs to disease transmission.

I. Five Points on the Economic Impact of the Pandemic and Recession on Women

1. The Pandemic's Impact on Female Labor Force Participation Rates and the Run-up before the Pandemic: *What Is the Correct Counterfactual?*

I will begin the elucidation of the impact of the pandemic on women with a discussion of female labor force participation rates. The existence of a sharp and unparalleled, in recent history, run-up in participation prior to the pandemic will influence the choice of a hypothetical to understand the impact of the pandemic.

The claim that the female labor force participation rate fell during the pandemic is not incorrect. But the implication that female labor force participation plummeted from a much higher level before the pandemic to one that was extremely low during the pandemic is highly mistaken.

First off, the 57 percent figure often mentioned in the media is for all women age 16 and older. Although that is a customary way of expressing the data and is done for historical consistency, it is not a meaningful age group to use. However, even using the age 16 and older group, the statistic for women fell by only 1 percentage point from April 2019, when it was 57.1 percent, to April 2021, when it was 56.1 percent. The same statistic fell by 1.4 percentage points for men, from 69.0 percent to 67.6 percent.[14]

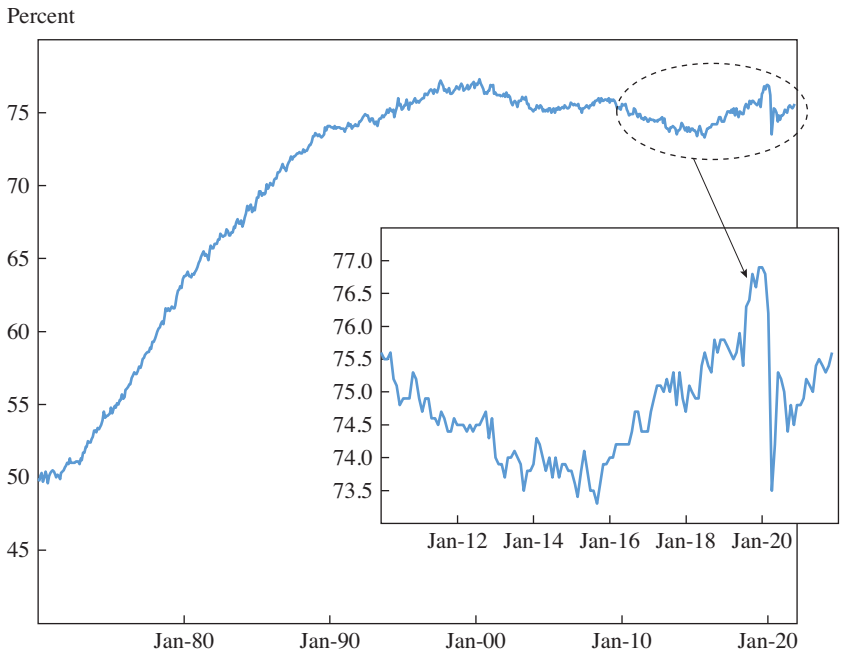
The reason that recent participation rates for women take us back many decades is that women's participation rates have not changed much during the past thirty years, and for some demographic groups, they actually decreased.[15] But men's participation rates have fallen almost every year since at least the 1960s.

13. I am not claiming that race and ethnicity mattered any more in the pandemic recession than in any previous one.

14. Data are seasonally adjusted from the Federal Reserve Bank of St. Louis, Economic Research Division (FRED), https://fred.stlouisfed.org, LNS11300002 series for women and LNS11300001 for men.

15. For data on the female labor force in general, see Goldin and Mitchell (2017). For information on the increase in female participation for those age 55 and older, see Goldin and Katz (2018).

Figure 1. Female Labor Force Participation Rate, January 1970 to November 2021



Source: Bureau of Labor Statistics, series LNS11300062, accessed March 2022.

Note: Seasonally adjusted, civilian labor force participation rate (expressed as a percentage), women age 25–54.

Labor force participation rates make more sense for a group of workers who are post-school, preretirement, say 25 to 54 years old, as depicted in figure 1 for the period since 1970. The labor force participation rate for women expressed in this manner shrank a bit from 75.5 percent in April 2019 to 75.1 percent in April 2021, comparing data for the same months both before and after the pandemic began, using the approximate month of the many media reports on the statistics. Expressed that way, the rate declined by just 0.4 percentage point. The rate for men, using the same age group, fell by 1.3 percentage points, considerably more. Comparing, say, November 2018 to November 2021 (the latest month for which data exist at the time of this writing) gives no change for women and a 0.7 percentage point decrease for men. The reason for using 2018 for this calculation rather than 2019 concerns a critically important run-up in women's labor force participation.[16]

16. Using 2019 gives a decrease of 1.0 percentage point for women and the same for men.

As is clear in figure 1, the January and February 2020 labor force participation rates for women are distinct outliers across a long period (see the enlarged portion of figure 1). The figure for January 2020 is 76.9 percent. Out of the 383 monthly numbers from January 1990 to November 2021, just ten exceed the January 2020 figure of 76.9 percent, and seven equal that figure, including the rate for December 2020. The February 2020 figure is 76.8 percent.[17]

The increase in female labor force participation rates in the mid-2010s was a return to a more normal era after the apparently delayed response of the female labor force to the Great Recession. By around September 2019, female labor force participation rates were approaching their level from before the Great Recession. Then came a boom in women's entry to the workforce.

Female labor force participation rates soared from late fall 2019 to early winter 2020, when the economy had exceptionally low unemployment. We may never know whether that increase in women's participation would have persisted in the absence of the pandemic or whether it would have ended up being another transitory blip.

We can, however, discern who entered the labor force in the period of run-up and what happened to the recent entrants during the pandemic. If those who entered largely remained in during the pandemic, then the increase might have been sustained. But if those who entered largely dropped out in the next several months, then the run-up consisted primarily of women who were marginally attached to the labor force. Comparisons with January 2020 would give an overstatement of the hypothetical impact of the pandemic in the absence of the run-up.

The answer, we will see, is that a substantial fraction of those who entered in the period immediately preceding the pandemic left during it. It is to be expected that recent entrants are, on average, less attached. But this group was large and somewhat less attached than in more ordinary times.

Consider, first, a simple descriptive summary of who entered just before the pandemic. According to the monthly Current Population Survey (CPS), not seasonally adjusted, the increase in participation among all women 20 to 54 years old from April 2019 to December 2019 was 1.86 percentage

17. The failure of the aggregate participation rate to return to its prepandemic levels is explored in Cooper and others (2021), but that paper emphasizes the aging of the population since November 2017 in the decreased aggregate participation rate rather than the run-up in participation by women, even though the run-up is evident in the authors' data.

points.[18] But among those age 20–29 who were not college graduates and had a child age 0–4, it was 6.32 percentage points. Therefore, the increase in labor force participation in the second half of 2019 was greatest among less-educated, young women with young children.[19]

To explore further, I use the longitudinal feature of the monthly CPS to understand the demographic characteristics of the women who entered the labor force just before the pandemic began and who among them remained in the labor force during the pandemic. The answer is that the new entrants were distinctive in the ways just described. In addition, they left the labor force during the pandemic at far greater rates than those who had not recently entered. It seems plausible, therefore, that the January and February 2020 figures are anomalous.

Each of the CPS respondents takes part in the CPS rotation, during which they are interviewed for four straight months and then for another four months after an eight-month hiatus. I first find all women age 20–54 who entered the labor force any time from April 2019 to February 2020. Each must have been interviewed at least twice in that period and been observed out of the labor force and then in. In addition, because they must have been last observed to have been in the labor force on or before February 2020, they would likely have been in the labor force at the start of the pandemic, in March 2020. Each of these women must also have had at least one observation in the pandemic period to determine if she remained in the labor force or left during the pandemic. The collection of the women who meet these conditions is labeled group 1.

As a control, I next identify women age 20–54 who were always in the labor force when interviewed between April 2019 and February 2020. They must have had at least one observation in the prepandemic period and another during or after March 2020 to observe their pandemic experience. That collective is labeled group 2.

Group 2 women, therefore, are always in the labor force when observed after April 2019 but before March 2020, whereas group 1 women enter the labor force at some point before the pandemic begins. Both groups are observed before and after the pandemic.

18. I use age 20 as the lower bound here and going forward because I would like to include parents with very young children. Beginning with 25-year-olds would exclude many parents of infants and toddlers.

19. This is not to say that college graduate women did not also have increased participation. It was just greater for other groups. Among college graduate women age 25–34 with a child younger than age 5, the increase from December to April was 3.69 percentage points, but it was 5.59 percentage points among those without a college degree.

Table 1. Characteristics of Women in the Labor Force

	Women age 20–54	
	Group 1 Entered the labor force from April 2019 to February 2020 and remained in to March 2020	*Group 2 Always in the labor force when observed from April 2019 to February 2020*
College graduates	0.326	0.452
With no children	0.518	0.478
With children under age 5	0.198	0.154
With children age 5–14	0.166	0.216
Ages 20–29	0.401	0.287
Left labor force March 2020 to last month observed	0.427	0.120
Number of observations	1,045	21,534

Source: CPS Monthly.

Note: Given the sample design, the initial interview could have occurred from April 2019 to January 2020, and the last interview from July 2020 to April 2021. As an example, consider an individual who began her CPS interviews in May 2019 (month 1) when she was out of the labor force. In July she entered the labor force and remained in at her August interview. We see her again eight months later in May 2020 (her CPS month 5) during the pandemic, and she would remain in the rotation until August 2020. She would be included in group 1 because she began out of the labor force but entered before the pandemic. She can then be observed after the pandemic. Group 2 women are always in the labor force when they are observed in the prepandemic period. See online appendix note 2 for details on the construction.

The results of the exercise, given in table 1, show that group 1 and group 2 were, not surprisingly, rather different demographically. The women of group 1 were less educated, were younger, and had more young children. Of real importance, 43 percent of group 1 left the labor force at least once after March 2020, whereas just 12 percent of group 2 did.[20]

An important implication of these findings is that the impact of the pandemic measured as a simple comparison of employment or labor force participation in any month after March 2020 with the same statistic in January or February 2020 will produce a larger estimate than one that differences from that month in, say, 2018. The simple difference leads to estimates that are overstated relative to an ideal counterfactual.

20. Recent entrants would be expected to be less attached to the labor force than the more established ones. I created a placebo pandemic in March 2018 and constructed two groups equivalent to those just described and for an equal set of months. Among the placebo group 1 women, 38 percent left the labor force at least once, but just 9 percent of placebo group 2 did. The impact is smaller than in the treatment period. The big difference is the relative size of the recent entrants.

In addition, the simple comparison or difference of a pandemic month with one in early 2020 just before the pandemic conflates seasonal changes with the pandemic impact.[21] In what follows, I will make comparisons with the same month or season in a prepandemic period, before the run-up in female labor supply, with that during the pandemic.

I will demonstrate that the comparison with winter 2019–2020 yields larger estimates of the labor force participation and "at work" rate declines for women in every season, but not for men, and that the estimates are larger for the less educated. Although I use months or seasons in 2018 as a standard, there are very few differences when 2019 is used as the reference year.[22]

2. Impact of the Pandemic and the Recession on Women Relative to Men: *Was the Pandemic Recession a She-cession?*

There are several ways to estimate the impact of the pandemic, each constructing a counterfactual concerning what employment or labor force participation would have been in the absence of COVID-19. The most reasonable estimate of what a group would have been doing in the absence of the pandemic is what the group had been doing in the same month in a previous, more normal, year. As just explained, using a period that preceded the run-up in women's labor force participation eliminates a potentially spurious component, and differencing on the same month removes seasonality.[23]

I have grouped months by season and performed a simple difference from the season in question to the same season in 2018–2019. I use the seven seasons from spring 2020 to fall 2021. Figure 2 shows the results for women and men 20 to 54 years old by education (college graduate versus not) for employment: "at work" in panel A and labor force participation in panel B.[24]

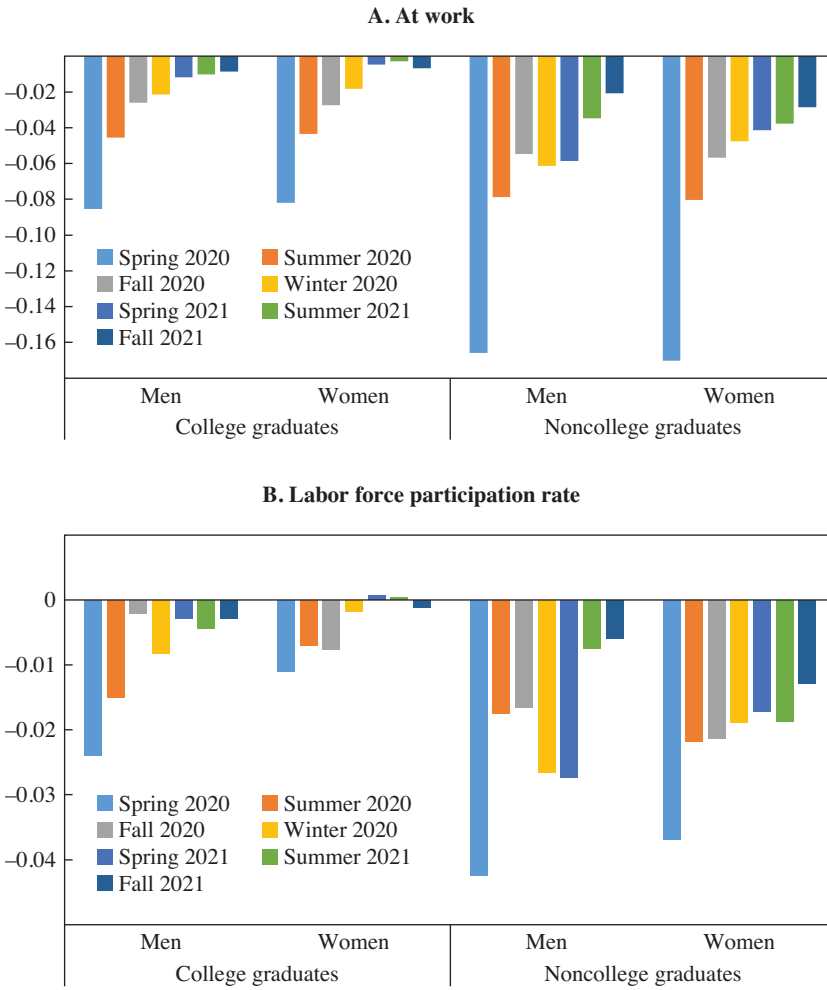
21. The seasonality issue arises from the use of the CPS micro data to look at subgroups, instead of using the Bureau of Labor Statistics' seasonally adjusted data or constructing a seasonally adjusted series from the micro data.

22. Because winter includes January and February of the following year, I will refer to December 2019 to February 2020 as winter 2019–2020. The full reference years will be termed 2018–2019 (March 2018 to February 2019) and 2019–2020 (March 2019 to February 2020).

23. Counterfactual estimates that difference each month show no relative increase by gender during the summer, suggesting seasonality is a major factor. Price and Wasserman (2022) explore why data for college graduate women show summer seasonality in employment and the possibility that K-12 teachers hired on twelve-month salaries report they are not at work in the summer.

24. The results are not much different if the age 25–54 group is used. I employ the age 20–54 group for consistency with later results that add the impact of young children.

Figure 2. "At Work" and Labor Force Participation Rate Changes, by Season, Education, and Sex



Source: CPS Monthly (Flood and others 2021).

Note: For men and women age 20–54. The values given for difference from the same season in 2018. Seasons are defined as spring (March–May), summer (June–August), fall (September–November), and winter (December–February), except spring 2020 excludes March 2020 and so does the comparison season in 2018. See online appendix table 1 for sample summary statistics.

Panel A excludes individuals who stated that they had a job but were not at work that week. That category is often high during the summer when many workers take vacation, and it could also indicate a parental or medical leave, but at the outset of the recession it was the work status given by many who were furloughed and did not know if they would be rehired. The base numbers for at work in 2018–2019 are useful for gauging the importance of the changes. In general, male college graduates have "at work" rates around 0.90 whereas female college graduates have rates around 0.80. Summer rates are lower for both: 0.87 for men and 0.74 for women. For the group who are not college graduates, the rates are around 0.78 for men and 0.64 for women. The notes to online appendix figure 1 provide "at work" base numbers for each of the seasons and groups shown here.

The fraction of the population at work excludes those who were out of the labor force or unemployed or who stated they had a job but were not at work during the CPS week. The latter category generally includes workers on short-term leave or vacation, but the group increased substantially during the pandemic. Thus, the most conservative estimate, and one that has become conventional in work on the pandemic, excludes this group from the "at work" population.

In absolute levels, the fraction at work declined significantly in spring 2020 for all groups but it declined considerably more for the less educated. The fraction at work decreased by a bit more than 8 percentage points for both male and female college graduates but by about 17 percentage points for the noncollege graduate group. The decrease had lessened a year later in spring 2021 when it was −1 percentage point for male college graduates and just −0.5 percentage points for female college graduates. It was approximately −6 for men and −4 percentage points for women in the noncollege graduate group.

Because men's "at work" base is larger than women's, the relative decrease was larger for women.[25] For college graduates, the magnitudes relative to the base levels are −9.45 percent for men and −10.35 percent for women for spring 2020 and −1.31 percent for men and just −0.61 percent for women in spring 2021. For the noncollege graduate group the relative magnitudes are larger: −21.2 percent for men and −26.5 percent for women in spring 2020 and −7.2 percent for men and −6.5 percent for women in spring 2021.

The (absolute) decrease in the fraction at work for college-graduate men was approximately the same as for college-graduate women 20 to 54 years old for all seasons and was actually lower for women in the most recent three seasons shown. Noncollege-graduate women experienced a slightly more negative impact in some seasons relative to men but not in others.

There does not seem to be a large difference in "at work" losses by gender in absolute terms using the counterfactual provided in figure 2, panel A. Rather, the large differences are by education. College graduates experienced half the decline in the fraction at work than did those with less education from spring 2020 to fall 2020. From winter 2020 to fall 2021, the less-educated group recovered somewhat less and had deficits in "at work" rates of around 4 percentage points.

25. The base "at work" rates by season in 2018–2019 for male college graduates are around 90 percent, whereas for female college graduates they are around 80 percent. Summer "at work" rates are lower for both: 87 percent for males and 74 percent for females. For the group who are not college graduates, the rate is around 78 percent for males and 64 percent for females for all seasons.

Employing, instead, the method that differences from winter 2019–2020 (see online appendix figure 1, panel A) produces larger changes in at work for women in every season. There are fewer differences among men, as should be expected since the substantial run-up occurred among women, not men. Furthermore, the absolute declines in at work for women are considerably higher in the summer, especially for college-graduate women, who were more likely than those without a college degree to have had paid vacation days.

Using the difference from winter 2019–2020 adds a whopping −6 percentage points to the estimates for summer 2020 and 2021, since the seasonal effect is not eliminated by this form of differencing. Seasonality is also apparent for the more-educated men, for the same reason. In this case, using the difference from winter 2019–2020 adds −3 percentage points for summer 2020 and 2021. Among the less educated, differences are apparent only for women for whom the difference is −1 to −2 percentage points. The main point is that differencing from winter 2019–2020 produces considerably larger estimates of the decrease in at work and also for labor force participation (see online appendix figure 1, panel B).

The difference in the two methods is essentially the "at work" number for each season in 2018–2019 minus that for winter 2019–2020. Since at work in winter 2019–2020 was relatively high, the number is negative. For less-educated women the value is −1.34 percentage points, averaged across all seasons. The value for more-educated women is even more negative when the summer months are included and is −1.11 percentage points excluding them. These are substantial differences and amount to about a 20 percent greater decline among less-educated women relative to the method that differences by the same season. I should note that using 2019–2020 as the reference year, rather than 2018–2019, does not produce substantially different results (see online appendix figure 2, panels A and B).[26]

But even using the differencing method from winter 2019–2020, which probably overstates the economic impact of the pandemic and incorrectly credits seasonal changes to it, the most apparent disparity is still by education. For women, the decrease in at work for the noncollege graduate group is almost twice what it is for college graduates, disregarding the summer months. Similar differences exist by education for men.

Differences for labor force participation rates, computed identically to those for at work, are also much smaller by gender than are the differences

26. The largest difference for less-educated women between the 2019–2020 and 2018–2019 reference years is 1.7 percentage points in the summer months. The average difference is just 0.6 percentage points.

by education (figure 2, panel B). For college graduates, labor force participation rates by spring 2021 were about the same as they were in 2018, and that is true for both men and women. In fact, men had slightly larger decreases than did women. For the noncollege graduate group, decreases were considerably greater, and the differences between men and women are not large except in the most recent seasons shown. In fall 2021, the latest season for which we currently have data, women were behind by 1.3 percentage points relative to 2018 and men were behind by 0.6 percentage points. Earlier in the pandemic and through spring 2021, the college graduate group was far more shielded than those without a college degree.

As was the case for the "at work" differences, labor force participation rate declines using the difference from winter 2019–2020 are considerably larger for every season for women but not for men (compare figure 2, panel B with online appendix figure 1, panel B). The differences for some seasons are quite large and reflect the seasonality issue raised earlier.

There are several takeaways. One is that education produced a big divide in the pandemic, gender less so. Another is that the counterfactual used changes the answers substantially. Many researchers who have differenced from January or February 2020 (winter 2019–2020) probably did not realize the potential biases in doing so, particularly regarding the issue of seasonality.[27]

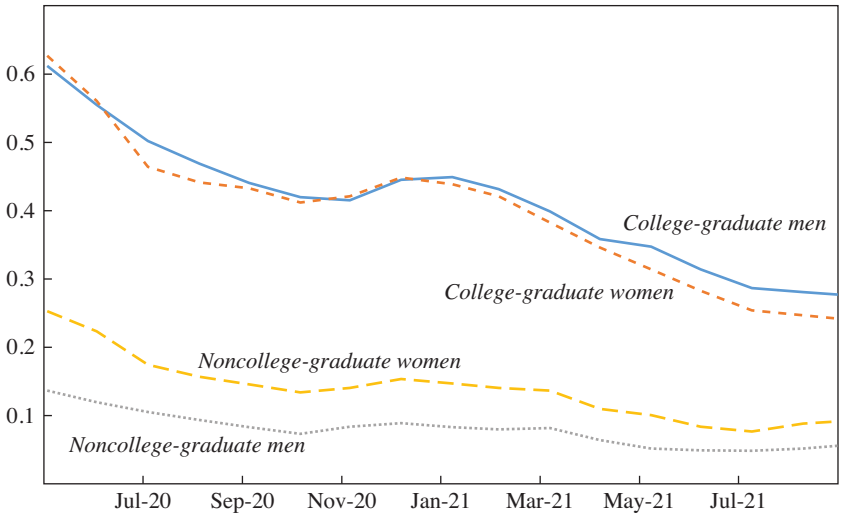
Education inoculated workers from the economic impact of the pandemic long before the vaccine was developed. The reason can be seen in figure 3, which graphs the answers to a CPS question, first asked in May 2020, about whether workers teleworked or worked for pay from home at any time during the previous four weeks "because of the coronavirus pandemic."[28]

Working remotely aided workers by enabling them to have safer jobs that could be done from home, and it enabled their firms and institutions to continue in operation. About 62 percent of employed college-graduate women and college-graduate men in May 2020 were working from home

27. Many papers have reported changes in employment and labor force rates using the difference from January or February 2020 or using one as a reference month. These include Couch, Fairlie, and Xu (2022), although the authors also difference by month; Furman, Kearney, and Powell (2021); Bauer, Estep, and Yee (2021); and Bauer and others (2021). Hansen, Sabia, and Schaller (2022), in research that uses the Safegraph data, give a time series that differences from January 2020 as motivation. Luengo-Prado (2021) generally differences from February 2020 but sometimes by month in the previous year.

28. Dingel and Neiman (2020), at the start of the pandemic, and before the CPS question was asked, produced estimates regarding which occupations could probably be done remotely.

Figure 3. Fraction of Employed Men and Women Who Worked Remotely, by Education



Source: CPS Monthly (Flood and others 2021).

Note: The full question asked is: "At any time in the last four weeks, did you telework or work at home for pay because of the coronavirus pandemic?" (The question was asked of people 16 years or older who were employed at the time of the survey; this figure includes men and women age 25–54.) See online appendix figure 3 for the full series by education level for men and women combined from BLS tabulations from May 2020 to January 2022.

at least part of the time due to COVID-19. But among those without a college degree, just 25.3 percent of employed women and 13.7 percent of employed men were working from home in May 2020 due to the virus. Because the CPS question was whether work was done remotely specifically because of COVID-19, the total fraction working at home would have been greater.[29]

By fall 2020, according to the CPS question, about 42 percent of the college graduate group continued to work from home, whereas around 13 percent of the noncollege-graduate women and just 7 percent of noncollege-graduate men did. In September 2021, around a quarter of college-graduate women worked remotely at least part of the time, but less than 10 percent of the noncollege-graduate women did and about only one in twenty of the noncollege-graduate men did.

29. To approximate an estimate that includes remote work in the absence of the pandemic, one can add to the CPS data an estimate of the usual fraction who worked remotely using American Time Use Survey (ATUS) data for 2017 and 2018. That would add about 10 percentage points (8 percentage points) to the numbers for college-graduate men (women), and 2 percentage points (3 percentage points) for noncollege-graduate men (women).

Interestingly, the fraction working remotely due to COVID-19 shot up in January 2022 for all education groups because of the Omicron variant (see online appendix figure 3). At that time, 29 percent of all college graduates (men and women) were working from home due to COVID-19, which was an increase of 33 percent from the previous month, December 2021.

It was C. Nicole Mason, president and chief executive of the Institute for Women's Policy Research, who first used the moniker "she-cession," a wordplay on the "man-cession" nickname for the 2008 recession.[30] There are many reasons why this recession was bound to be different.

State mandates at the start of the pandemic shut down or limited the density of many in-person services, such as those provided by restaurants, bars, hair salons, and retail stores. The travel and hospitality industries had greatly reduced demand. Jobs in these service sectors had seemed more immune to past cyclical downturns and other vagaries of the economy, such as the China trade shock and automation, than had those in goods-producing sectors. Prior to March 2020, women were more than 60 percent of employees in these sectors and occupations (see online appendix note 1: "Occupations Coded as 'Service'"). Men were in the more cyclically sensitive industries, such as manufacturing and construction.[31]

But the pandemic produced both a he-cession and a she-cession. Relative to previous recessions, women have been harder hit. But the largest differences in pandemic effects on employment are found between education groups rather than between genders within educational groups.

The other reason the pandemic should have impacted women more than men is that the care sectors—K-12 schooling, day care, and elder care—were also shut down or made remote. With limited in-person schooling, childcare, and elder care, mothers and adult daughters largely filled the gap. Economic changes due to the pandemic in labor force participation and at work rates are those we can measure. But they are only part of the time demands on parents, especially women. I turn now to the home front.

3. Childcare Hours for Working Parents in the Pandemic's First Year: *Did Childcare Hours Double Initially and How Did Hours Change by Fall/Winter 2020–2021?*

When schools closed, day care centers were shuttered, nannies were sent packing, and grandparents were sequestered, childcare demands on parents

30. "We should go ahead and call this a 'shecession'" (Gupta 2020, quoting Mason). https://www.nytimes.com/2020/05/09/us/unemployment-coronavirus-women.html.

31. See, for example, Albanesi and Kim (2021).

soared. The same is true for adult children, disproportionately daughters, who cared for elderly parents when home health care aides could no longer work and when residents were removed from toxic nursing homes.

A problem in assessing just how much caring time increased is that many of the facts are not yet known and may never be known for a large enough group. The American Time Use Survey (ATUS) stopped during the pandemic, and although it commenced again in May 2020, the sample size from the early pandemic period is small.

The good news is that several surveys were executed in the United States and Europe during the pandemic, and some were continuing surveys that had existed before the pandemic. But samples vary regarding whether both parents were present, whether they worked, if they worked remotely, the ages of the children, and what gets included in childcare hours.

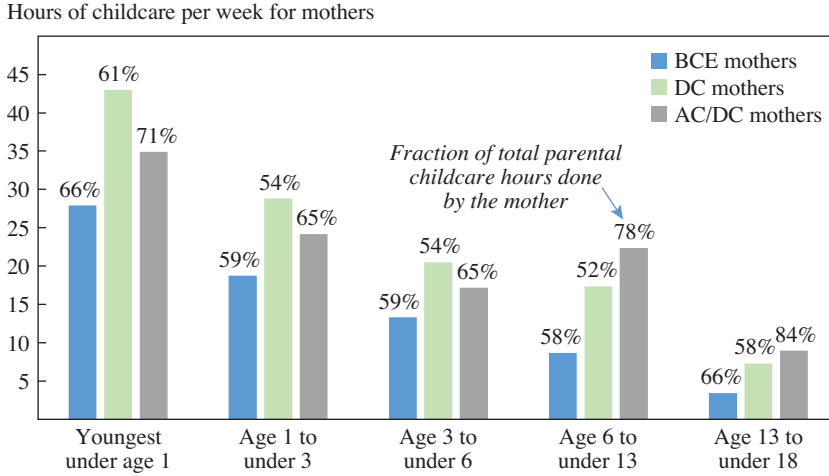
By piecing together the evidence from the ATUS, available before the pandemic, with surveys in the United States and Britain during the pandemic for which sufficient consistency exists, it does appear that childcare hours doubled in families of working parents in the immediate aftermath of the lockdown in spring 2020. Not only did the childcare hours of mothers increase but also the share of the total done by (custodial) fathers, at least for a while.

To create a consistent sample before and during the pandemic, I consider college-graduate, employed (different sex) parents who have at least one resident child younger than 18 years and who live together. The reason I chose college graduates is that each would have had a high probability of being able to work remotely during the pandemic, and much of the survey evidence concerns those who worked at home. In addition, the college graduate group would have had a high probability of maintaining their jobs during the pandemic, and many of the surveys consider only those who were employed.

I have used the ATUS to compute prepandemic childcare hours of (custodial) mothers and fathers by the age of their youngest child regardless of the number of children.[32] The first bar in figure 4, for each of the child ages, gives (weekly) childcare hours of the mother before the COVID-19 era (BCE), as gleaned from the ATUS. The years 2010 to 2019 are used to obtain a large enough sample to stratify by couple characteristics. The fraction of total parental childcare hours for mothers is given above each bar. Before March 2020, college-graduate employed mothers did around

32. I use the main activity and do not add secondary childcare time (for those less than age 13, the group included in the question), which allocates all main activity time to childcare if it contains any secondary childcare time.

Figure 4. Childcare Hours of College-Graduate, Employed Women with College-Graduate Employed Husbands, by the Age of Their Youngest Child



Sources: ATUS for 2010–2019; Andrew and others (2020).

Note: BCE = before the COVID-19 era (2010–2019); DC = during COVID-19 (spring 2020); AC/DC = after COVID-19 but during COVID-19 era (fall 2020 through winter 2021 = September 2020 to February 2021). BCE hours come from a sample of women in the ATUS who were currently employed, college graduates, with at least one child less than 18 years old and a husband who was also a college graduate and currently employed. Daily childcare amounts are multiplied by seven. Children are "household" children (children who live in the household, independent of whether one or more of the parents live there), and childcare includes all types of care, including educational. All days of the week are included. Numbers above the bars are the percentage of total parental childcare hours provided by the mother. DC hours are estimated by increasing BCE hours by 1.54 for mothers and 1.9 for fathers, which are the ratios from Andrew and others (2020), and then adding an additional four hours per week (per parent) when the youngest child is age 6 to under 13 and two hours when the youngest is age 13 to under 18. AC/DC hours for the couple are an average of BCE and DC hours, but fathers are given only BCE childcare hours under the assumption that they are back at work full-time. Mothers are assumed to be doing the rest of the childcare. The average is one-quarter the difference between BCE and DC hours for children under age 6, but three-quarters for those age 6 to under 18.

60 percent of total childcare hours (not including housework, laundry, and food prep and cleanup, the addition of which would increase the fraction since women in the ATUS did around 70 percent).

The second bar in figure 4, for each of the child ages, denotes childcare hours of mothers during COVID-19 (DC) in spring 2020, when almost 90 percent of US school-age children were in school remotely and most childcare facilities were shuttered. The data come in part from Andrew and others (2020) because that study extended time budget results from the prepandemic era and provides changes in hours with the pandemic. The authors find that childcare hours for (custodial) fathers increased by about 1.9 times and by 1.54 times for mothers.

Many aspects of Andrew and others' (2020) sample families are consistent with the ones I am using from the ATUS, but some are not. In addition, the ATUS requests the actual time period, but the survey used in Andrew and others (2020) allocates a task to a block of time, independent of the actual time spent at it. Increasing the ATUS childcare time by the fractions in the study resulted in time use that seemed low given home schooling reported in other surveys. In consequence, I added four hours per week to two-parent families with a youngest child age 6–12 and two hours for those with a youngest child age 13–17 for consistency with other surveys.[33] That produced the data for the DC period.

Since many of these households had both parents (college graduates) at home full time, parental sharing increased. Consequently, the fraction of childcare performed by mothers fell, even as total parental childcare hours doubled and as the childcare hours of mothers increased by around 1.7 times (more than 1.54 because of the additional child education hours).

In fall 2020, we moved into the after COVID-19 but during COVID-19 era (AC/DC). Draconian pandemic restrictions were partially lifted, and some offices allowed workers to return, others demanded they do. Day care centers were allowed to open in most states, although some had already gone out of business. Schools in many large districts did not fully open, and some that opened had in-person instruction only intermittently.

The third bar, for each child age group, provides estimates of childcare hours in fall 2020 through winter 2021, the AC/DC period. The bars contain underlying data whose levels are somewhere between the values shown in the BCE and DC bars. The assumptions are that total childcare levels decreased for the youngest children more than for school-age children since day care was generally open for more time than were elementary, middle, and high schools. Custodial fathers are assumed to have returned to their prepandemic levels of childcare; mothers absorb the entire difference.[34]

33. Pre-March 2020 childcare hours in Andrews and others (2020), created to be consistent with their COVID-19 sample, are high, and thus the ratio of post- to pre- may not be high enough. Even with the added hours, the totals are far less than those in Adams-Prassl and others (2020), who give a whopping 6.1 hours per workday for mothers (and 4.8 hours for fathers) of children age 4–15. But Carlson, Petts, and Pepin (2021) interviewed 1,025 US parents with at least one biological child and asked about changes in housework and childcare after COVID-19 restrictions were implemented. They found (using CPS weights) that the vast majority said that they did about the same amount of childcare, with just 27 percent of mothers and 36 percent of fathers saying they did more. Sevilla and Smith (2020) interviewed 2,782 parents in the United Kingdom, again with a before-and-after question, and found large increases in childcare time. A survey by Krentz and others (2020) asked about the combination of childcare and household tasks and found a doubling after the pandemic for working parents but provided no information on child ages.

34. See figure 4 notes for the assumptions used.

There was probably no net gain for working mothers in the move from the DC world (spring 2020) to the AC/DC world in fall 2020 through winter 2021. What they gained from partial and often sporadic school and day care openings, they likely lost from less parental help at home as more men than women went back to their offices and work sites or worked more intensively on their jobs from home. In consequence, mothers' total childcare hours remained about the same, but their share of the total increased.

The statement by many that parents of young and school-age children doubled their childcare time overnight in spring 2020 is likely correct.[35] Mothers greatly increased their housework and care hours, and even if their childcare hours may not have fully doubled, the sheer number of hours became an enormous burden, especially for those with full-time jobs. Custodial fathers' childcare time also increased and probably more than doubled, having started out at lower levels than mothers'. There was greater sharing among parents as time burdens increased for both.

The shift back to offices and job sites left mothers in fall 2020/winter 2021 with a larger fraction of childcare time even as the total number of their childcare hours remained the same. Much of the frustration expressed by mothers in the AC/DC era concerned the fact that schools in many parts of the United States had not yet reopened or had only partly opened, but many fathers had returned to their workplaces or had ramped up their remote job hours to full time.[36]

I mentioned earlier that the ATUS resumed in May 2020 and is currently available to December 2020 (the ATUS pandemic period). Because sample sizes are small, I have used all cohabiting individuals with a child younger than age 18, classified in the child groups given in figure 4. The ATUS also provides data for a prepandemic period (January 2019 to February 2020) using the same weights as in the pandemic months.

Using data for the non-summer pandemic months shows that relative to the prepandemic period, educational care of children increased and accounted for all of the increased childcare time.[37] For example, women with a spouse and a youngest child age 3–5 stated they spent 16.0 hours per week caring for their children before the pandemic and 18.2 hours per week

35. Meakin (2020), based on data from a Boston Consulting Group (BCG) survey conducted from March 20 to April 3, 2020, referring to both childcare and housework time.

36. Pabilonia and Vernon (2022), using the 2020 ATUS detailed diary data, find that by fall 2020 only mothers put in more childcare hours when both parents worked for pay from home although at the start of the pandemic both did.

37. Sample sizes are small and few cells exceed fifty observations. I have used data for the same months in 2019 as in 2020 (May and September to December).

during the pandemic. Educational childcare, included in the total, increased from 2.2 to 5.9 hours. Therefore, additional educational time exceeded the total childcare hours increase. Similarly, for the age 6–12 group, the total increase was from 9.1 to 10.4, and education time increased by more, from 1.0 to 3.7 hours. Fathers during the pandemic reported increased childcare time with their newborns by about 25 percent (from 12.6 to 15.9 hours).

But the total increase in childcare time among all ATUS parents is far less than in other studies. Whereas total parental time doubled using the assumptions underlying figure 4, the pandemic 2020 ATUS increase in primary childcare time is puny. In fact, there are groups for which childcare time apparently decreases. The reason appears to concern the distinction between primary and secondary childcare time in the ATUS.[38] Secondary childcare time allocates all time in a non-childcare task to childcare if any time is spent watching or helping children 12 years and younger.

With more parents at home, secondary childcare time increased even as primary childcare time did not. For employed mothers with a child younger than age 13, secondary childcare time increased from 4.1 hours per weekday (for non-summer months) in 2019 to 6.1 in 2020. The increase for employed fathers was from 3.0 to 3.7 hours. Primary care for these parents actually decreased a small amount as more parents were working at home. But secondary childcare time is difficult to compare across time when remote work changes and hard to equate with its primary analogue.[39] More data from the ATUS are needed to make better sense of how the increase in remote work has had an impact on childcare and will continue to have an impact on all time uses.

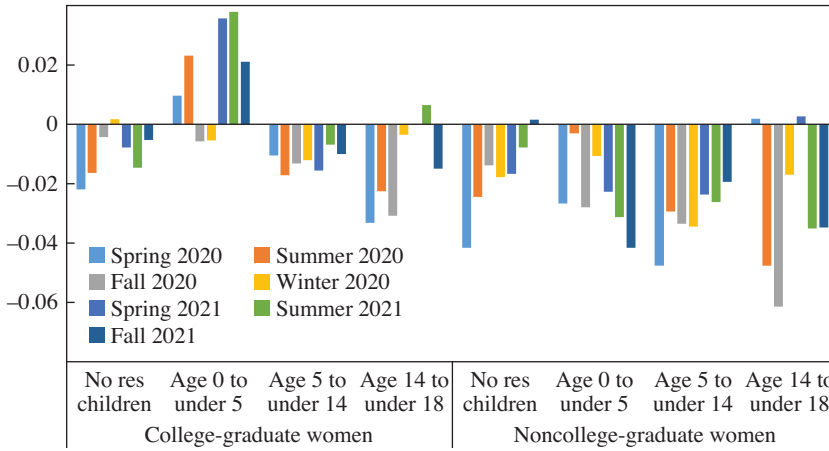
4. Impact of the Pandemic on Women: *How Did Employment and Labor Force Participation Change for Mothers by the Age of Their Youngest Child and for Women without Residential Children?*

Relative to other national economic crises, the pandemic produced more stresses and setbacks for women. The reasons are several. Women were disproportionately in the more vulnerable jobs and their time as caregivers for children and others greatly expanded. How their labor force participation rates and fraction at work varied by the age of their youngest child is given in figures 5 and 6. The data demonstrate, once again, that those with

38. Bauer, Estep, and Yee (2021) compute primary and secondary childcare time for the 2020 ATUS.

39. Sample sizes are small (two hundred respondents). Secondary childcare expands on weekends to include almost all non-sleep and personal care time, demonstrating that parents, especially mothers, often double task.

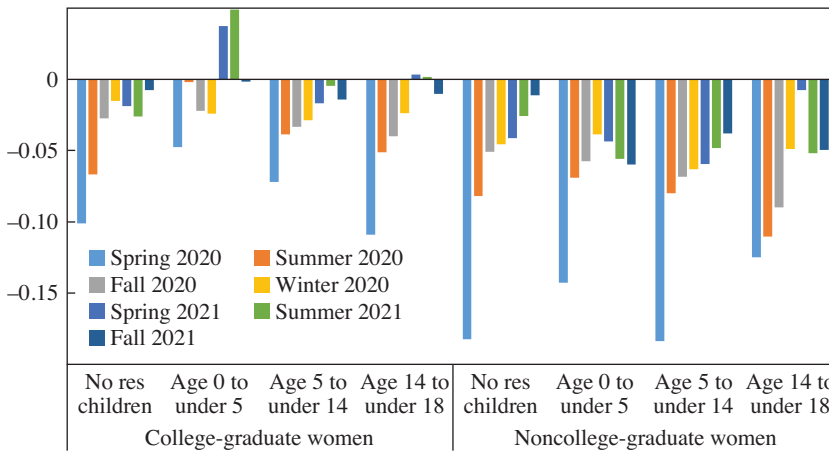
Figure 5. Labor Force Participation Rate Changes by Education Level and Age of Youngest Child



Source: CPS Monthly (Flood and others 2021).

Note: For women age 20–54. Seasons are defined as spring (March–May), summer (June–August), fall (September–November), and winter (December–February), except spring 2020 excludes March 2020 and so does the comparison season in 2018. "No res children" refers to no residential children under 30 years old; a woman with no residential children could have children. Information for women with residential children age 18–29 years was omitted from the figure. See online appendix table 1 for sample summary statistics.

Figure 6. "At Work" Rate Changes for Women, by Education Level and Age of Youngest Child



Source: CPS Monthly (Flood and others 2021).

Note: For women age 20–54. Seasons are defined as spring (March–May), summer (June–August), fall (September–November), and winter (December–February), except spring 2020 excludes March 2020 and so does the comparison season in 2018. "No res children" refers to no residential children less than 30 years old; a woman with no residential children could have children. Information for women with residential children age 18–29 is omitted from the figure. See online appendix table 1 for sample summary statistics.

a college degree weathered the storm far better than those who lacked one, in part because they could (and did) work remotely. Note that, as before, differences are taken from a particular season in 2018–2019 to a season during the pandemic to avoid using a period during the run-up in women's labor force participation as well as to adjust for seasonality.[40]

Beginning first with changes in labor force participation rates given in figure 5, decreases were relatively small for the college graduate group, except for those with teenage children at the start of the pandemic. In fact, college-graduate women with infants and toddlers had increased participation rates relative to 2018–2019, particularly after winter 2020. Working at home may have opened doors and options for them. Note that there was even an immediate increase, which may have resulted from having both parents working at home (the pandemic ATUS shows that fathers with infants increased their childcare time more than did fathers with older children). Perhaps new mothers who would have left the labor force in ordinary times decided to stay.

Not so for similar mothers without a college degree for whom work at home was often not an option, for them or for their spouses, and new jobs that could be done with added childcare demands were less available. In fact, noncollege-graduate women with a youngest child under age 5 had decreased labor force participation rates in summer and fall 2021, while the college graduate group experienced the opposite.

For all noncollege-graduate women with children under age 18, labor force rates remained 2 to 4 percentage points (or more) below their 2018–2019 levels even by fall 2021, the last season for which we currently have data.[41]

Labor force participation is a bellwether of future employment, whereas unemployment is a measure of current harm and income loss. Leaving the labor force means greater costs to reenter employment. That is why the possibility that women's participation had substantially decreased during the pandemic has been so concerning and may be why the McKinsey-LeanIn survey responses (Thomas and others 2020) were accorded great attention

40. As noted before in the discussion of the robustness of the results by sex and education, the use of 2019–2020 as the reference year does not greatly change the results by the age of the youngest child. See online appendix figure 4, panels A and B.

41. Other researchers have also found that the role of children differed for college-graduate and noncollege-graduate women and that occupation was more important for the noncollege graduate group (Alon and others 2021; Luengo-Prado 2021). Although not shown, Black college-graduate women fared relatively well, but Black women without a college degree had larger declines than their white counterparts.

and filled headlines with dire predictions. But the evidence is clear. Women by and large did not leave employment at the extensive margin.

I'll demonstrate the point quite simply. The labor force participation rate for all college-graduate women, 25 to 34 years old, in fall 2021 was 85.5 percent, and it had been 85.4 percent in fall 2018. It had actually increased. For those with children, the figures are 78.2 percent in fall 2021 and 77.2 percent in fall 2018. Once again, the rate increased.

Using, instead, the hardest-hit months in 2020 (April and May), the labor force participation rate was 82.7 percent; it was 84 percent for the same months in 2018. It fell by 1.3 percentage points. Using the 2018 base, this implies that one in sixty-five college-graduate women in the 25–34 age group had exited the labor force. For those with a child, the participation rate was 73.6 percent in 2020 (again just April and May) and 74.3 percent in 2018 (same months). In that case about one in a hundred had left.

For the entire noncollege graduate group 25 to 34 years old, the numbers are 69.7 percent for fall 2021 and 70.8 percent for fall 2018, a decline of just 1.1 percentage points. For the worst months of the pandemic, spring 2020 (April and May), the labor force participation rate was 66.1 percent; it was 70.9 percent for spring 2018, a decline of 4.8 percentage points—a lot more than for the college graduate group. For those with children, the labor force participation rate was 61.5 percent in spring 2020 but 65.9 percent in spring 2018, a decline of 4.4 percentage points.

Even the largest decrease, for the hardest-hit group—noncollege graduates in spring 2020—meant that about one in fifteen had exited the labor force. These statistics, while lamentable, are a far cry from the one in four and the one in three figures cited in the McKinsey-LeanIn survey (Thomas and others 2020) and broadcast widely and persistently by the media.

The McKinsey-LeanIn survey, although substantial in size, consisted of employees who occupied six job levels: executives, senior VPs, VPs, senior managers, managers, and entry-level office and corporate employees, such as customer service reps (Thomas and others 2020, 59). These are occupational titles of a highly educated, occupationally elite group. The one in four and one in three numbers, in this widely cited survey, seem even more ludicrous.

Many reports, including the survey just cited, mentioned a scaling back in hours by women, especially mothers. In addition to a reduction in labor force participation, the pandemic could also have had an impact on the intensive margin through a change in hours. Among college-graduate women 25 to 34 years old who were at work a year or less before March 2020 and who remained at work a year later, about 32 percent reported

Table 2. Share of Group with Changes in Hours Worked

	Pandemic phase		
	(1) Pre-pre	*(2) Pre-pan*	*(3) Pan-pan*
	Decrease in hours		
College graduate			
Age 25–34	0.267	0.322	0.246
Age 25–54	0.280	0.326	0.255
Age 25–54 with child under age 13	0.257	0.297	0.228
Not college graduate			
Age 25–34	0.269	0.328	0.246
Age 25–54	0.271	0.333	0.249
Age 25–54 with child under age 13	0.265	0.326	0.266
	Increase in hours		
College graduate			
Age 25–34	0.257	0.227	0.307
Age 25–54	0.272	0.244	0.309
Age 25–54 with child under age 13	0.262	0.243	0.320
Not college graduate			
Age 25–34	0.306	0.267	0.337
Age 25–54	0.294	0.252	0.340
Age 25–54, with child under age 13	0.330	0.272	0.342

Source: CPS Monthly.

Note: All were at work at the start and end of phase. "Pre-pre" indicates that the period is entirely before March 2020; "pre-pan" indicates that the period traverses March 2020; "pan-pan" means that the period is entirely after February 2020. The numbers in the table are the fraction with a decrease or increase in hours, with no change being the omitted group. The sample is restricted to those who were employed in both t and $t + 1$. Actual hours on all jobs are differenced between start month and end month a year later if the respondent was at work in both periods. See online appendix note 3 for details on the data construction.

some reduction in hours (see table 2, column 2). But 27 percent of those in the same demographic group reported a reduction in hours before the pandemic (column 1), also across a year for which they were at work at the start and the end (therefore 41 percent reported no change in hours). Not surprisingly, workers report changes in their hours in non-pandemic times as well as in pandemic times.

Therefore, 5 percent (32 – 27 percent) or one in twenty college-graduate women 25 to 34 years old who were at work both before and after the pandemic decreased their hours at the start of the pandemic relative to an ordinary period of an equivalent length. For those without a college degree, it was 6 percent (33 – 27 percent) or one in seventeen. It should be noted that the fraction of the college graduate group who increased hours during the

pandemic is almost identical to what it was before the pandemic among those at work. I discuss the construction of the sample used in table 2.

The belief that more-educated women would drop out of the labor force persisted despite evidence to the contrary. In large part, the notion remained because mothers and other caregivers were stressed and increasingly made that known to reporters, who were also often mothers and were stressed. But the reporters and their sources were strained and frazzled because they *didn't* drop out of the workforce. Employed women who were helping to educate their children and working adult daughters who were caring for their parents were stressed because they were in the labor force, not because they had left.[42] The real story of women during the pandemic is that they remained in the labor force. They stayed on their jobs, as much as they could, and persevered.

Findings for "at work" rates can be seen in figure 6, which gives results using a counterfactual similar to that on labor force participation. Deficits were substantial at the start of the pandemic for all women, even those without young children. But they were, as before, especially large for noncollege-graduate women, often double those of the college graduate group. The college graduate group managed to make up considerable ground across the pandemic whereas noncollege-graduate women often lost ground (as occurred for those with children age 0–4) or made little headway (as occurred for the group with children age 14–17).

For the most recent months, college-graduate women with children had an employment deficit relative to 2018 of only around −1 percentage point, but the noncollege graduate group with children had a deficit of around −5 percentage points. For much of the pandemic period, however, even the noncollege graduate group without residential children had about the same employment deficits as those with young and school-age children.[43] But by fall 2021, the group without residential children (or with children older than 17) had largely returned to work, whereas noncollege-graduate women with younger children continued to lag.

Some of the lag in the "at work" numbers can be seen in the continued high unemployment rates for the less-educated group. Among women 25 to 34 years old without a college degree, 6.9 percent were still unemployed

42. Garcia and Cowan (2022) find that school closures had little impact on whether parents worked at all but did have an impact on the intensive margin among lower-educated parents.

43. I use the term "residential" children rather than "dependent" children because they are identified as children of a mother or father because they are living in the household.

in fall 2021. Even though that is substantially lower than the 10.2 percent they experienced in fall 2020, it is higher than the 5.6 percent figure from fall 2018.

An important point for both college-graduate and noncollege-graduate women is that even women without residential children fared poorly for much of the pandemic and that was particularly true for less-educated women through winter 2020. But like the previous analysis, the big divide in employment recovery was less about children and more about education and the ability of women to have jobs that were protected in more ways than one. I will unpack this further in the next section.

5. Putting It All Together: *What Were the Separate Impacts of Children, Education, Occupation, Race, and Marital Status on Employment Transitions before, at the Inception of, and during the Pandemic?*

To explore the joint roles of the various forces already discussed, such as parenthood, occupations, education, and race, in the pandemic period, I created an additional extract using the longitudinal feature of the monthly CPS. The data track individuals from exactly one year to the next for the same month. The age group used is 20 to 54 years old to include more women with young children.

Due to the panel structure of the CPS in which individuals are interviewed for four consecutive months and an additional four months after an eight-month hiatus, the individuals in the extract would have been interviewed in month 1 in year t and month 5 in year $t + 1$, or month 2 in year t and month 6 in year $t + 1$, and so forth using CPS-month notation. In creating the data set in this fashion, I observe the same individual in the same month but a year later. Some will traverse the period before the pandemic, some will begin before the pandemic but be interviewed again after it began, and others will have data from within the pandemic period. These are the same data that were used in table 2 on changes in hours of work.

The full period I explore begins in January 2018 and extends to November 2021. There are three possible pandemic phases:

Within the prepandemic period, that is, prepandemic to prepandemic (termed pre-pre; 43 percent of observations), with t from January 2018 to February 2019, so that an individual can be tracked across January 2019 to February 2020, $(t + 1)$, just before the pandemic began.

Prepandemic to pandemic period (termed pre-pan; 35 percent of observations), with t from March 2019 to February 2020, so that an individual can be tracked from March 2020 to February 2021, $(t +1)$, beginning before the pandemic and ending during the pandemic.

Within the pandemic period (termed pan-pan; 23 percent of observations), with t from March 2020 to November 2020, so that an individual can be tracked from March 2021 to November 2021, $(t + 1)$, during the pandemic. The last month and year of the data used are November 2021.

The resulting extract produces the dependent variable in the equation, $\Delta y^{m}_{i,t}$, which is the change in either "at work" or labor force participation for individual i in month m for year t relative to the same month a year later in $(t + 1)$. It is defined here as a (1, 0) variable for which the individual is either (in, in) = 1 or (in, out) = 0. Thus, I restrict the sample to begin with individuals at work or in the labor force. I will discuss only the "at work" results since those on the labor force are similar in kind and smaller in magnitude. The setup ensures that information on the individual's prior occupation is available even if the person exited employment or the labor force by $(t + 1)$.

The variables included in the equation are all indicator variables and are intended to gauge the separate strength over the pandemic periods of a set of variables that covary, such as education, race, parental status, age, and occupation:

$$\begin{aligned}\left(y^{m}_{i,t} - y^{m}_{i,t+1}\right) &= \Delta y^{m}_{i,t} \\ &= \alpha + \sum_{\varphi=1}^{2}\left[\beta \times I(\varphi)\right] + \gamma I^{i,t}(C) + \delta I^{i}(E) + \theta I^{i,t}(O) + \rho I^{i}(R) \\ &\quad + \mu I^{i,t}(M) + \eta I^{i}(X) + \lambda + \kappa + \epsilon_{i,t}.\end{aligned}$$

The value of β gives the role of each of the pandemic phases (φ) relative to the pre-pre group. Main effects are given by indicators for the youngest child's age in five bins (C), where "no residential children" (grouped with "residential children older than 29") is omitted; an indicator for college graduate (E); an indicator for pandemic-impacted occupations (O), mainly in the service sector and defined in online appendix note 1; indicators for Black and Hispanic (R); and an indicator for unmarried or unpartnered (M). All main effects (C, E, O, R, M) are interacted with the two pandemic phases pre-pan and pan-pan, as is the interaction between marital status and a youngest child under age 5. X is a set of indicators for the individual's age in five-year bins, λ is a set of year dummies, and κ is season dummies, defined in the table 3 notes (online appendix table 2 provides summary statistics).

The regression in table 3 includes only women (a pooled regression is later discussed). Table 3, column 1, contains the main effects and an interaction of the pandemic phases with college graduate. The interaction

Table 3. Annual Changes in "At Work" Rates for Women, January 2018 to November 2021

	Women at work in year t, *month* m			
	Main effects	*Plus child interactions*	*Plus race and occupation interactions*	*Plus marital interactions*
Respondent's age				
20–24	−0.0788***	−0.0788***	−0.0787***	−0.0849***
	(−23.21)	(−23.23)	(−23.19)	(−24.37)
25–29	−0.0182***	−0.0182***	−0.0180***	−0.0219***
	(−5.92)	(−5.91)	(−5.86)	(−7.04)
30–34	−0.0154***	−0.0153***	−0.0152***	−0.0180***
	(−5.02)	(−4.99)	(−4.97)	(−5.86)
35–39	0.00991**	0.0100***	0.0104***	0.00775*
	(3.28)	(3.32)	(3.43)	(2.55)
40–44	0.00887**	0.00895**	0.00941**	0.00798**
	(3.02)	(3.05)	(3.21)	(2.72)
45–49	0.00935***	0.00944***	0.00968***	0.00915**
	(3.34)	(3.37)	(3.46)	(3.27)
Youngest child's age				
0–4 years	−0.0277***	−0.0194***	−0.0197***	−0.0137***
	(−11.40)	(−5.37)	(−5.46)	(−3.65)
5–13 years	−0.00870***	−0.00631*	−0.00547	−0.00141
	(−3.94)	(−1.97)	(−1.71)	(−0.43)
14–17 years	0.00815**	0.000809	0.00215	0.00551
	(2.67)	(0.18)	(0.48)	(1.20)
18–29 years	0.00654*	0.00533	0.00721	0.00984*
	(2.05)	(1.14)	(1.55)	(2.09)
College graduate	0.0173***	0.0169***	0.0242***	0.0256***
	(6.98)	(6.80)	(9.43)	(9.92)
Black	−0.0205***	−0.0205***	−0.00920*	−0.0129***
	(−8.77)	(−8.76)	(−2.56)	(−3.54)
Hispanic	−0.0311***	−0.0311***	−0.0298***	−0.0309***
	(−14.61)	(−14.61)	(−9.23)	(−9.56)
Service occupation	−0.0562***	−0.0562***	−0.0253***	−0.0263***
	(−25.45)	(−25.46)	(−7.71)	(−7.98)
Start year is 2018	−0.0298***	−0.0298***	−0.0298***	−0.0298***
	(−4.69)	(−4.69)	(−4.69)	(−4.69)
Start year is 2019	−0.0334***	−0.0335***	−0.0333***	−0.0334***
	(−8.05)	(−8.06)	(−8.03)	(−8.05)
Spring	−0.0185***	−0.0186***	−0.0186***	−0.0186***
	(−6.64)	(−6.65)	(−6.67)	(−6.68)
Summer	−0.0236***	−0.0236***	−0.0237***	−0.0239***
	(−8.39)	(−8.38)	(−8.44)	(−8.51)
Fall	0.00825**	0.00826**	0.00828**	0.00819**
	(2.98)	(2.98)	(2.99)	(2.96)
Prepandemic to pandemic (pre-pan)	−0.103***	−0.0994***	−0.0719***	−0.0714***
	(−23.14)	(−20.34)	(−13.70)	(−12.21)

(*continued on next page*)

Table 3. Annual Changes in "At Work" Rates for Women, January 2018 to November 2021 (*Continued*)

	Women at work in year t*, month* m			
	Main effects	*Plus child interactions*	*Plus race and occupation interactions*	*Plus marital interactions*
Pandemic to pandemic (pan-pan)	−0.0463*** (−6.47)	−0.0480*** (−6.42)	−0.0425*** (−5.47)	−0.0418*** (−5.02)
Pre-pan × college	0.0576*** (16.10)	0.0582*** (16.22)	0.0398*** (10.49)	0.0385*** (10.08)
Pan-pan × college	0.00925* (2.28)	0.00973* (2.39)	0.00633 (1.48)	0.00512 (1.19)
Pre-pan × children under age 5		−0.0218*** (−4.15)	−0.0211*** (−4.01)	−0.0130* (−2.14)
Pan-pan × children under age 5		−0.00374 (−0.63)	−0.00365 (−0.61)	0.00356 (0.52)
Pre-pan × children age 5–13		−0.00859 (−1.87)	−0.0105* (−2.28)	−0.0103* (−2.14)
Pan-pan × children age 5–13		0.00237 (0.45)	0.00142 (0.27)	0.00143 (0.26)
Pre-pan × children age 14–18		0.0141* (2.15)	0.0107 (1.64)	0.0104 (1.56)
Pan-pan × children age 14–18		0.0107 (1.42)	0.00950 (1.27)	0.00918 (1.20)
Pre-pan × children age 18–30		−0.00174 (−0.26)	−0.00419 (−0.62)	−0.00458 (−0.66)
Pan-pan × children age 18–30		0.00789 (1.03)	0.00647 (0.85)	0.00595 (0.77)
Pre-pan × Black			−0.0243*** (−4.57)	−0.0229*** (−4.24)
Pan-pan × Black			−0.0133* (−2.16)	−0.0118 (−1.89)
Pre-pan × Hispanic			−0.0103* (−2.13)	−0.00969* (−1.99)
Pan-pan × Hispanic			0.00859 (1.55)	0.00920 (1.66)
Pre-pan × service occupation			−0.0786*** (−15.92)	−0.0781*** (−15.79)
Pan-pan × service occupation			−0.0164** (−2.79)	−0.0155** (−2.64)
No spouse				0.0147*** (3.78)
Pre-pan × no spouse				0.000470 (0.09)
Pan-pan × no spouse				0.000856 (0.18)
Pre-pan × no spouse × children under age 5				−0.0304*** (−3.70)

Table 3. Annual Changes in "At Work" Rates for Women, January 2018 to November 2021 (*Continued*)

	Women at work in year t*, month* m			
	Main effects	*Plus child interactions*	*Plus race and occupation interactions*	*Plus marital interactions*
Pan-pan × no spouse × children under age 5				–0.0311** (–2.95)
Constant	0.963*** (166.22)	0.962*** (163.60)	0.951*** (159.65)	0.944*** (153.71)
Number of observations	174,226	174,226	174,226	174,226

Source: CPS Monthly (Flood and others 2021).

Notes: Dependent variable (0, 1) indicates whether a respondent who was at work in year *t*, month *m*, was also at work in the same month in year $t + 1$. All observations begin at work. The period is divided into three phases: before the pandemic (pre-pre, beginning January 2018 to February 2019); prepandemic to pandemic (pre-pan, beginning March 2019 to February 2020); and pandemic to pandemic (pan-pan, beginning March 2020 to November 2020). The last month of the data is November 2021. "No spouse" includes individuals not currently married or partnered. Omitted variables: respondents age 50–54; no children resident or resident child is older than 29 years; not a college graduate; white; start year 2020; winter; phase 1. Service occupations are those that were most negatively affected by state mandates that restricted the operation of businesses or were in industries that were; they are listed in the online appendix note 1. See online appendix note 3 for details of the construction.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

with college graduate reinforces the results from the cross section data (discussed above) showing that more-educated women were substantially inoculated from job loss. The shift into the pandemic (pre-pan) decreased the fraction at work among all women by –0.103 but college-graduate women experienced less than half that reduction (–0.103 + 0.0576).

Interactions of the pandemic phases with the age of the youngest child are added in column 2. Relative to the omitted group, only those with the youngest and high school–age children have an additional impact from the shift into the pandemic. The finding that children have little added impact as the pandemic spread should not be surprising given that women with no residential children (either no children or no younger children at home) had as large or larger employment shortfalls throughout the pandemic, with the exception of the most recent months (see figure 6). Although the finding is strongest for the less educated, it also holds for the college graduates. All women were greatly impacted by the pandemic.

Column 3 adds interactions with the race and ethnic groups and also with service occupations, a set of occupations and those in particular industries that were often shut down at the start of the pandemic and have continued

to experience negative impacts from the pandemic. About 18 percent of all employed women in the pre-pre sample (2018 and early 2019) were in these occupations, as were 27 percent of the noncollege graduate group and 33 percent of Black noncollege-graduate women. These were important occupations especially for less-educated women and for Black women.

Women in these service occupations and industries have always experienced greater employment instability, as can be seen from the main effect. But they experienced an even greater negative shock in the transition to the pandemic. Similarly, Black and Hispanic women have always experienced greater instability of employment than white women, but the change going into the pandemic was an additional effect, separate from having been employed in one of the more-severely impacted occupations, having children of various ages, and so forth.[44] Column 4 adds interactions with the variable "no spouse," for which the interactions identify being a single mother. The sign and magnitude demonstrate that the youngest children had a large impact because many had single moms.

The main takeaways are illustrated in figure 7. The descent into the pandemic period had a great impact on all women, and the fraction who were at work fell by 7.2 percentage points, although a college degree cushioned the fall by 4 percentage points. Having a youngest child under age 5 produced a negative impact of 2.1 percentage points and having a youngest child age 5–13 reduced at work by 1 percentage point. The largest changes were experienced by those who began in one of the service occupations, as they suffered an additional decline of 7.9 percentage points.[45] Given all the included variables, Black women experienced an additional 2.4 percentage points decline and Hispanic women a 1 percentage point decline.[46]

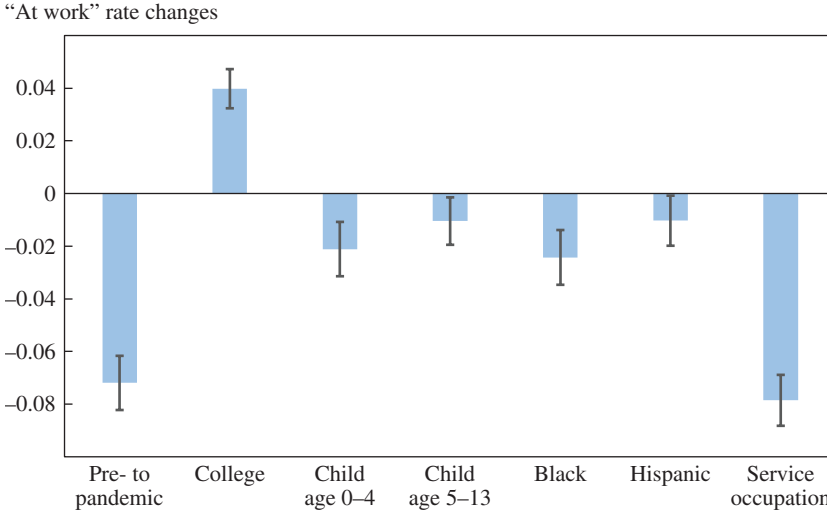
There is little additional effect during the pandemic period (pan-pan) among those who began that phase at work. All women (using column 3 in table 3) had a decrease of 4.3 percentage points. There was no added cushion for the college graduates. Workers in service occupations had an additional 1.6 percentage point decrease in being at work during this period.

44. I also ran a version (not shown) with the race variables but not the occupation variable and then compared the results with column 3, showing that race and occupation are orthogonal in this analysis.

45. Including a full set of two-digit occupation dummies does not reduce the protective role of being a college graduate in the pre-pan phase.

46. Luengo-Prado (2021) also finds that occupation was more important than children in mothers' employment.

Figure 7. Impact of the Pandemic on "At Work" Rates of Women



Source: Author's calculations.

Note: Bars indicate the strength of the interaction of the variable with "Pre- to pandemic," which refers to whether individuals employed from March 2019 to February 2020 ("pre-") remained at work or not a year later (March 2020 to February 2021). Child variables refer to the age of the youngest child in the household. "Service occupation" is a group of particular occupations, as well as those in some industries, that were generally shut down or had limited density of operation at the start of the pandemic by state mandates (see online appendix note 1). Since all individuals were employed at the start of the period considered, the occupation refers to that in the prepandemic period. Confidence intervals at 95 percent are shown.

The US Census Household Pulse Survey data provide complementary evidence about why Black women in the pandemic recession experienced an impact beyond the variables included in the regression. The data demonstrate that the health of Black women and of those in their households were key factors.[47] Respondents were asked the most important reason they were not employed. Black women, 25 to 44 years old with children younger than age 18, were far less likely than similar white women to say they could not work because they were caring for children. But they were more likely to have been laid off or furloughed, and they were far more likely to have said they were sick with COVID-19 or caring for someone who was as reasons for nonemployment.

47. The survey data were designed to provide rapid evidence on the impact of COVID-19 on individuals. The micro data are used here for September 29 to October 11, 2021, and December 29, 2021, to January 10, 2022. See US Census Bureau, "Household Pulse Survey Public Use File (PUF)," https://www.census.gov/programs-surveys/household-pulse-survey/datasets.html.

Among those without children younger than age 18, 34 percent of Black women were either ill (with COVID-19 or another ailment), caring for someone, or feared getting ill at work, whereas 22 percent of white women gave these reasons. The data for surveys from December 2021 to January 2022 reveal the damage the Omicron variant has done, especially in the Black community. Among women 25 to 44 years old, 10 percent of white women, but 20 percent of Black women, said that COVID-19 had kept them out of work.

This paper began with the finding that the impact on "at work" and labor force rates for men and women was about equal in absolute value during the pandemic and that the largest differences were by education. I also noted that the impact of having a youngest child at particular ages varied by season and year of the pandemic and that race was an additional factor.

Because the economic impact of the pandemic was about the same by sex, combining men and women in an analysis similar to that from the equation used here with the full range of interactions does not add much to the discussion from the previous analysis, which used data from repeated cross sections (results are given in online appendix table 3). Both men and women had a decline of 6.1 percentage points in the fraction at work going into the pandemic. Given the covariates, women had an additional decrease in their fraction at work of 1.38 percentage points, but women with a college degree had a lesser impact by 1.39 percentage points—almost the same amount—than similar men (all college graduates experienced a 2.6 percentage point boost).

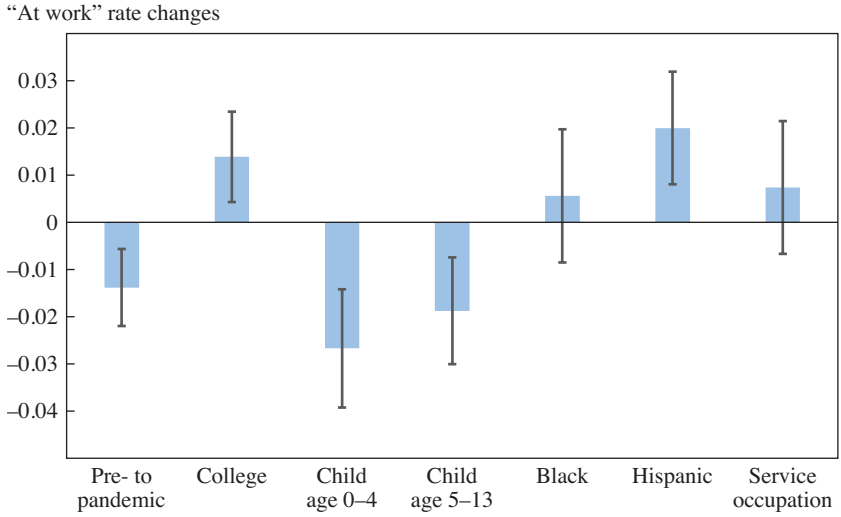
Thus, lower-educated women were the ones who did less well than their equivalently educated male counterparts. Differences by race and ethnicity were substantial in the pandemic phase, but differences by gender do not reveal much in addition. The same is true for the service occupations. Figure 8 summarizes the results.

What about the role of children? Women with children age 0 to 4 experienced an impact 2.7 percentage points greater in the pre-pan phase than did men with residential children of those ages, and those with children age 5–13 experienced an impact 1.9 percentage points greater.[48]

An additional word must be said about which men are in the comparison group. Many who have contributed to the literature on the impact of

48. Couch, Fairlie, and Xu (2022) find a similar effect for those with children age 0–4.

Figure 8. Impact of the Pandemic on "At Work" Rates of Women Relative to Men



Source: Author's calculations.
Note: The coefficient is the interaction of gender with the variable given and the pre-pan phase. Confidence intervals at 95 percent are shown.

the pandemic have looked at differences between men and women by the age of their youngest residential child. The list of papers is long, yet few have recognized that the CPS does not identify all fathers. Men (and women) who do not reside with children may still have and care for children. In fact, according to the US Census more than 30 percent of all children younger than 18 years who reside with a woman (generally their mother) do not live with their father.[49] But their fathers live somewhere. Thus, men without children in their household may be fathers, albeit noncustodial ones.

In discussing results from estimations that use child ages and interactions with gender, one must identify the individuals as residential parents or guardians. As such, the comparison is not to those without children but

49. See Hemez and Washington (2021), who give the fraction of children living with one parent.

rather to those without residential children. The findings are that women with a youngest child under age 5 had reduced "at work" rates in the year they went into the pandemic relative to men with residential children under age 5. Similarly for those with the youngest age 5–13, although the effects are considerably smaller.

II. Summing Up and Looking Ahead

There is no question that the pandemic had a greater impact on women than other recessions have done. It was a she-cession relative to other recessions and also relative to the January (or February) 2020 labor force and "at work" statistics. But gender differences month by month in employment outcomes, relative to prepandemic levels, are not large. The big differences are by education rather than gender, and that makes it more similar to previous recessions.

Mothers greatly increased their time spent in childcare during the pandemic, but custodial fathers did as well. Female labor force participation in the United States did not plummet to its lowest level since the late 1980s. Its growth had been anemic for some time relative to that experienced in comparable nations. With the exception of older women, the female labor force participation rate in the United States has not increased in three decades. It decreased during the pandemic but actually by less than it did for men.

Far more mothers and other women who are caregivers have been stressed, frustrated, and anxious because they did *not* leave their jobs than have been forced to exit the workforce or cut back on their hours. Black women who were not college graduates were hardest hit in terms of their employment and labor force participation.

As noted earlier, the fraction of women who had decreased their paid work hours in the prepandemic-to-pandemic phase is larger than in the other phases. But it is not much larger, and a substantial group increased their hours. In addition, the fraction with decreased hours was generally a bit smaller, not larger, for the group with children. Changes in the intensive margin were not substantial going into the pandemic (starting prepandemic and ending in the pandemic) relative to equivalent changes in a more normal period (ending prepandemic).

What accounts for the excessive statements in the popular press, even from veteran writers who know the territory well? One reason is that individual accounts reported in the news are from those experiencing the most adversity. Another is that surveys such as that done by McKinsey-LeanIn

(Thomas and others 2020) captured the stresses and frustrations of the moment rather than actions.

It is precisely the mothers who did not drop out who expressed the greatest anxieties about their future careers. Because these women still had jobs, they worried about their current productivity and whether they could do enough to merit the raise or promotion, make partner or tenure.[50] They have been torn between being a good parent and doing their own paid work, an issue that predated the pandemic but has been magnified. Finally, the CPS—the primary data source that I and many others have used—reveals nothing about what individuals do with their time spent not at work or their mental well-being.[51]

There is the possibility that we will emerge on the other side of darkness with benefits. The cost of workplace flexibility will probably fall as workers, firms, managers, clients, and customers use what they have learned during lockdown to work more effectively as we open up.[52] If a contract can be signed without the expensive trip to Tokyo or Beijing, parents of young children will benefit from less travel time and firms will profit from lower costs. Given the current division of household labor, mothers will reap the greatest returns and will be able to take on more lucrative positions that once required considerable time away from home and were outside the realm of possibility for many.

Work at home, Zoom meetings, telemedicine and teletherapy, workouts and more online may have taught us how to work and live more efficiently without travel, overnight stays, and in-person meetings. A reduction in the cost of temporal and geographic flexibility may be part of a silver lining to the pandemic for women.

Surveys from late spring 2021 found that the majority of workers did not want to return to the office or job site five days a week and would rather

50. Deryugina, Shurchkov, and Steans (2021) surveyed academics from May to July 2020 and showed that research time decreased for all parents but decreased more for mothers. Flaherty (2020) used Elsevier journal data for the early pandemic months to show that publications of women generally lagged those of men.

51. Zamarro and Prados (2021) use data from the USC Dornsife Center for Economic and Social Research on household division of labor and mental health measures among parents.

52. "Flexibility" is a multidimensional concept that involves both temporal and geographic flexibility. It often means the ability of workers to control their hours in terms of the number and the moment in time. It can also mean the ability of employees to work from home as well as in a different city.

continue working at home one or two days.[53] A recent Gallup poll (March 2022) found that just 9 percent of workers want to return in person full time, and the majority (59 percent) would like a hybrid model (of course, these are workers who can work remotely).[54] The Bureau of Labor Statistics, in February 2022, reported that of the 34.5 percent of establishments that increased telework for some or all employees, 60.2 percent expected to maintain increased telework when the pandemic recedes (BLS 2022). The Survey of Working Arrangements and Attitudes found, as of January 2022, that workers who could work from home believed their employer will have them work remotely around two days a week after the pandemic ends.[55] These possibilities should be a plus for those with caregiving demands.

Many corporate leaders have been bullish on workplace flexibility. Last year's headlines on the subject were almost universally positive (David 2021; Anders 2021).[56] More recent headlines show less certainty and concern over productivity and fairness in workplaces fractured by time and space. Workers not only want to work at home, they also want to work in places often geographically distant from their previous offices. Above all, they want to work hours they choose (Bindley and Cutter 2022).

We are two years into the pandemic, and we do not yet know how it will play out for women. If, in the new normal, men go to the office five days a week and women go to the office three days a week and work from home two days, women won't be part-timers in terms of hours, but they will be part-timers in terms of face time and time spent with colleagues in the office. Women will do the client-facing meetings on Zoom, and men will go to Zürich to close the deal. The new normal at work may increase female labor force participation in the short run, but, like its part-time hours equivalent, it may not come with the same bonuses, pay increases, and promotions.

53. Barrero, Bloom, and Davis (2021), using survey data, estimate that 20 percent of full workdays will be work from home after the pandemic ends, whereas 5 percent were before. They also estimate productivity boosts that will show up in conventional productivity measures and cost savings from less commuting that will not show up. Bloom and others (2015) measure productivity increases from telecommuting. Emanuel and Harrington (2021) demonstrate negative selection to telework but productivity boosts given negative selection. The last two papers mentioned concern call centers, which probably do not have much increased productivity and creativity from group interactions.

54. Ben Wigert, "The Future of Hybrid Work: 5 Key Questions Answered With Data," *Gallup Workplace* (March 15, 2022). https://www.gallup.com/workplace/390632/future-hybrid-work-key-questions-answered-data.aspx

55. Barrero, Bloom, and Davis (2021); see also WFH Research, "Data," https://wfhresearch.com/data/ for the most recent data.

56. "There are early signs that remote work can help level the playing field" (Thomas and others 2020, 51).

Until more workers take advantage of work flexibility in all its forms, women who take the amenity could pay a career price in the long run. They may not lose as much as when they worked part-time, and they may not lose as much as when they changed jobs and firms to enhance their work flexibility. All depends on whether the pandemic will soften the greediness of work by making flexible jobs more productive, reducing the premium from the greedier jobs, and lowering the penalty from the more flexible ones.[57]

In addition, the gains to women from added flexibility will depend on keeping schools and day care facilities open and making elder care safer. Even among the fortunate who could work from home during the worst of the pandemic, productivity appears to have been related to their parental status. Recent studies have shown that mothers with academic jobs issued fewer working papers during the pandemic and published fewer journal articles than did fathers with academic jobs and female colleagues without school-age children.[58] But the worst-hit women in terms of health concerns and job security have been those who could not work from home. They disproportionately served others in hospitals and grocery stores, had incomes and education levels that were low, and had children with the least access to remote learning technology. They won't gain as much from the new normal.

We know considerably more about what happened to workers and in workplaces during the past two years. We still know little about what will happen and what the shape of the new normal will be for women.

ACKNOWLEDGMENTS Discussants Stefania Albanesi and Jane Olmstead-Rumsey and *BPEA* coeditor Janice Eberly provided valuable comments on an earlier version. Sara Estep was an extraordinary fact checker and resource. Joy Wang gave outstanding assistance linking the CPS monthly data, and Yailin Navarro provided excellent research help. I thank Dev Patel and Larry Katz for extensive comments on a draft. An earlier version, "Assessing Five Statements about the Economic Impact of COVID-19 on Women," was written for the Societal Experts Action Network (SEAN) of the National Academies initiative presented on June 24, 2021, at the SEAN Women in the Work Force webinar.

57. By "greedy work," I mean work that pays more, implicitly, on an hourly basis when workers work long hours or work specific hours, such as in the evenings, on vacation, or on weekends. For more on "greedy work" and a historical perspective, see Goldin (2021).

58. See, for example, Deryugina, Shurchkov, and Steans (2021).

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): 1–27.

Albanesi, Stefania. 2021. "The Impact of the COVID-19 Recession on the US Labor Market: Occupation, Family, and Gender." Presented at the National Academies of Sciences, Engineering, and Medicine, Societal Experts Action Network (SEAN), COVID-19's Impact on Women's Work Experiences: Challenges and Opportunities Webinar, June 24. https://www.nber.org/sites/default/files/2021-06/SEAN_Albanesi.pdf.

Albanesi, Stefania, and Jiyeon Kim. 2021. "The Gendered Impact of the COVID-19 Recession on the US Labor Market." Working Paper 28505. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w28505/w28505.pdf.

Alon, Titan, Sena Coskun, Matthias Doepke, David Koll, and Michele Tertilt. 2021. "From Mancession to Shecession: Women's Employment in Regular and Pandemic Recessions." Working Paper 28632. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w28632/w28632.pdf.

Alon, Titan, Matthias Doepke, Jane Olmstead-Rumsey, and Michele Tertilt. 2020a. "The Impact of COVID-19 on Gender Equality." Working Paper 26947. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w26947/w26947.pdf.

Alon, Titan, Matthias Doepke, Jane Olmstead-Rumsey, and Michele Tertilt. 2020b. "This Time It's Different: The Role of Women's Employment in a Pandemic Recession." Working Paper 27660. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w27660/w27660.pdf.

Anders, George. 2021. "How the Pandemic Changed Us: Our Fastest Rising Priority Is Job Flexibility." LinkedIn News, April 14. https://www.linkedin.com/pulse/how-pandemic-changed-us-our-fastest-rising-priority-job-george-anders.

Andrew, Alison, Sarah Cattan, Monica Costa Dias, Christine Farquharson, Lucy Kraftman, Sonya Krutikova, Angus Phimister, and Almudena Sevilla. 2020. "How Are Mothers and Fathers Balancing Work and Family under Lockdown?" Briefing Note. London: Institute for Fiscal Studies. https://ifs.org.uk/publications/14860.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis. 2021. "Why Working from Home Will Stick." Working Paper 28731. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w28731/w28731.pdf.

Bauer, Lauren, Eliana Buckner, Sara Estep, Emily Moss, and Morgan Welch. 2021. *Ten Economic Facts on How Mothers Spend Their Time*. Washington: The Hamilton Project, Brookings Institution. https://www.brookings.edu/research/ten-economic-facts-on-how-mothers-spend-their-time/.

Bauer, Lauren, Sara Estep, and Winne Yee. 2021. "Time Waited for No Mom in 2020." Blog post, July 22, Brookings Institution. https://www.brookings.edu/blog/up-front/2021/07/22/time-waited-for-no-mom-in-2020/.

Bindley, Katherine, and Chip Cutter. 2022. "Workers Care More about Flexible Hours Than Remote Work." *Wall Street Journal*, January 25. https://www.wsj.com/articles/workers-care-more-about-flexible-hours-than-remote-work-11643112004.

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.

Bureau of Labor Statistics (BLS). 2022. "U.S. Business Response to the COVID-19 Pandemic Summary." News release. February 9. https://www.bls.gov/news.release/covid2.nr0.htm.

Carlson, Daniel L., Richard Petts, and Joanna R. Pepin. 2021. "Changes in US Parents' Domestic Labor during the Early Days of the COVID-19 Pandemic." *Sociological Inquiry*. https://doi.org/10.1111/soin.12459.

Connley, Courtney. 2021. "Women's Labor Force Participation Rate Hit a 33-Year Low in January, According to New Analysis." CNBC, February 8. https://www.cnbc.com/2021/02/08/womens-labor-force-participation-rate-hit-33-year-low-in-january-2021.html.

Cooper, Daniel H., Christopher L. Foote, María J. Luengo-Prado, and Giovanni P. Olivei. 2021. "Population Aging and the US Labor Force Participation Rate." Federal Reserve Bank of Boston. https://www.bostonfed.org/publications/current-policy-perspectives/2021/population-aging-and-the-us-labor-force-participation-rate.aspx.

Couch, Kenneth A., Robert W. Fairlie, and Huanan Xu. 2022. "The Evolving Impacts of the COVID-19 Pandemic on Gender Inequality in the US Labor Market: The COVID Motherhood Penalty." *Economic Inquiry* 60, no. 2: 485–507. https://doi.org/10.1111/ecin.13054.

David, Roger. 2021. "Flexibility: Every Corporation's Most Important Strength during the Pandemic." Blog post, Forbes Business Council, March 1. https://www.forbes.com/sites/forbesbusinesscouncil/2021/03/01/flexibility-every-corporations-most-important-strength-during-the-pandemic/?sh=619fea1e734f.

Derrington, Taletha J., Alison Huang, and Joseph P. Ferrie. 2021. "Life Course Effects of the Lanham Preschools: What the First Government Preschool Effort Can Tell Us About Universal Early Care and Education." Working Paper 29271. Cambridge Mass: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w29271/w29271.pdf.

Deryugina, Tatyana, Olga Shurchkov, and Jenna E. Steans. 2021. "COVID 19 Disruptions Disproportionately Affect Female Academics." Working Paper 28360. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w28360/w28360.pdf.

Dingel, Jonathan I., and Brent Neiman. 2020. "How Many Jobs Can Be Done at Home." Working Paper 26948. Cambridge, Mass.: National Bureau of Economic

Research. https://www.nber.org/system/files/working_papers/w26948/w26948.pdf.

Emanuel, Natalia, and Emma Harrington. 2021. "'Working' Remotely? Selection, Treatment, and Market Provision of Remote Work." Working Paper. https://scholar.harvard.edu/eharrington/publications/working-remotely-selection-treatment-and-market-provision-remote-work.

Flaherty, Colleen. 2020. "Women Are Falling Behind." *Inside Higher Ed*, October 20. https://www.insidehighered.com/news/2020/10/20/large-scale-study-backs-other-research-showing-relative-declines-womens-research.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, and Michael Westberry. 2021. *Integrated Public Use Microdata Series, Current Population Survey: Version 9.0* [dataset]. Minneapolis: IPUMS. https://doi.org/10.18128/D030.V9.0.

Forman, Tami. 2022. "Women's Workforce Participation Has Plummeted. Here's How to Reverse the Trend." *Fortune*, February 4. https://fortune.com/2022/02/04/womens-workforce-participation-has-plummetedheres-how-to-reverse-the-trend-pandemic-labor-shortage-diversity-economy-tami-forman/.

Furman, Jason, Melissa Schettini Kearney, and Wilson Powell. 2021. "The Role of Childcare Challenges in the US Jobs Market Recovery during the COVID-19 Pandemic." Working Paper 28934. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w28934/w28934.pdf.

Garcia, Kairon Shayne D., and Benjamin W. Cowan. 2022. "The Impact of School and Childcare Closures on Labor Market Outcomes during the COVID-19 Pandemic." Working Paper 29641. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w29641/w29641.pdf.

Goldin, Claudia. 1991. "The Role of World War II in the Rise of Women's Employment." *American Economic Review* 81, no. 4: 741–56.

Goldin, Claudia. 2021. *Career and Family: Women's Century-Long Journey toward Equity*. Princeton, NJ: Princeton University Press.

Goldin, Claudia, and Lawrence F. Katz. 2018. "Women Working Longer: Facts and Some Explanations." In *Women Working Longer*, edited by Claudia Goldin and Lawrence F. Katz. Chicago: University of Chicago Press.

Goldin, Claudia, and Joshua Mitchell. 2017. "The New Life Cycle of Women's Employment: Disappearing Humps, Sagging Middles, Expanding Tops." *Journal of Economic Perspectives* 31, no. 1: 161–82.

Goldin, Claudia, and Claudia Olivetti. 2013. "Shocking Labor Supply: A Reassessment of the Role of World War II on Women's Labor Supply." *American Economic Review* 103, no. 3: 257–62.

Gupta, Alisha Haridasani. 2020. "Why Some Women Call This Recession a 'Shecession.'" New York Times, May 9. https://www.nytimes.com/2020/05/09/us/unemployment-coronavirus-women.html.

Hansen, Benjamin, Joseph J. Sabia, and Jessamyn Schaller. 2022. "Schools, Job Flexibility, and Married Women's Labor Supply: Evidence from the COVID-19

Pandemic." Working Paper 29660. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w29660/w29660.pdf.

Heggeness, Misty. 2020. "Why Is Mommy So Stressed? Estimating the Immediate Impact of the COVID-19 Shock on Parental Attachment to the Labor Market and the Double-Bind of Mothers." Working Paper 33. Federal Reserve Bank of Minneapolis. https://www.minneapolisfed.org/research/institute-working-papers/why-is-mommy-so-stressed-estimating-the-immediate-impact-of-the-covid-19-shock-on-parental-attachment-to-the-labor-market-and-the-double-bind-of-mothers.

Hemez, Paul, and Chanell Washington. 2021. "Percentage and Number of Children Living with Two Parents Has Dropped since 1968." US Census Bureau. https://www.census.gov/library/stories/2021/04/number-of-children-living-only-with-their-mothers-has-doubled-in-past-50-years.html.

Krentz, Matt, Emily Kos, Anna Green, and Jennifer Garcia-Alonso. 2020. "Easing the COVID-19 Burden on Working Parents." Boston Consulting Group. https://www.bcg.com/publications/2020/helping-working-parents-ease-the-burden-of-covid-19.

Luengo-Prado, María J. 2021. "COVID-19 and the Labor Market Outcomes for Prime-Aged Women." Federal Reserve Bank of Boston. https://www.bostonfed.org/publications/current-policy-perspectives/2021/covid-19-and-the-labor-market-outcomes-for-prime-aged-women.aspx.

Meakin, Lucy. 2020. "Parents' Chores and Child Care Almost Double during Pandemic." *Bloomberg Business*, May 21. https://www.bloomberg.com/news/articles/2020-05-21/coronavirus-almost-doubles-the-housework-for-working-parents.

Miller, Claire Cain. 2021. "The Pandemic Created a Child-Care Crisis. Mothers Bore the Burden." *New York Times*, May 17. https://www.nytimes.com/interactive/2021/05/17/upshot/women-workforce-employment-covid.html.

Pabilonia, Sabrina Wulff, and Victoria Vernon. 2022. "Who Is Doing the Chores and Childcare in Dual-Earner Couples during the COVID-19 Era of Working from Home?" Discussion Paper 15118. Bonn: Institute of Labor Economics. https://www.iza.org/publications/dp/15118/who-is-doing-the-chores-and-childcare-in-dual-earner-couples-during-the-covid-19-era-of-working-from-home.

Price, Brendan M., and Melanie Wasserman. 2022. "Summer Drop in Female Employment." Working Paper, UCLA Anderson School. https://static1.squarespace.com/static/57bdef2ecd0f687f9fbfec30/t/6297c8a712dcb120bb35bbce/1654114473519/PriceWasserman-SummerDrop.pdf.

Sevilla, Almudena, and Sarah Smith. 2020. "Baby Steps: The Gender Division of Childcare during the COVID-19 Pandemic." Discussion Paper 13302. Bonn: Institute of Labor Economics. https://docs.iza.org/dp13302.pdf.

Smart, Tim. 2021. "COVID-19 Job Market Wreaks Havoc on Black Women," *US News and World Report*. April 15. https://www.usnews.com/news/economy/articles/2021-04-15/black-women-suffering-the-most-from-covid-19-job-market-disruption.

Thomas, Rachel, Marianne Copper, Gina Cardazone, Kate Urban, Ali Bohrer, Madison Long, and others. 2020. *Women in the Workplace*. McKinsey and Company and LeanIn.Org. https://wiw-report.s3.amazonaws.com/Women_in_the_Workplace_2020.pdf.

Zamarro, Gema, and María J. Prados. 2021. "Gender Differences in Couples' Division of Childcare, Work and Mental Health during COVID-19." *Review of Economics of the Household* 19, no. 1: 11–40.

Comments and Discussion

COMMENT BY

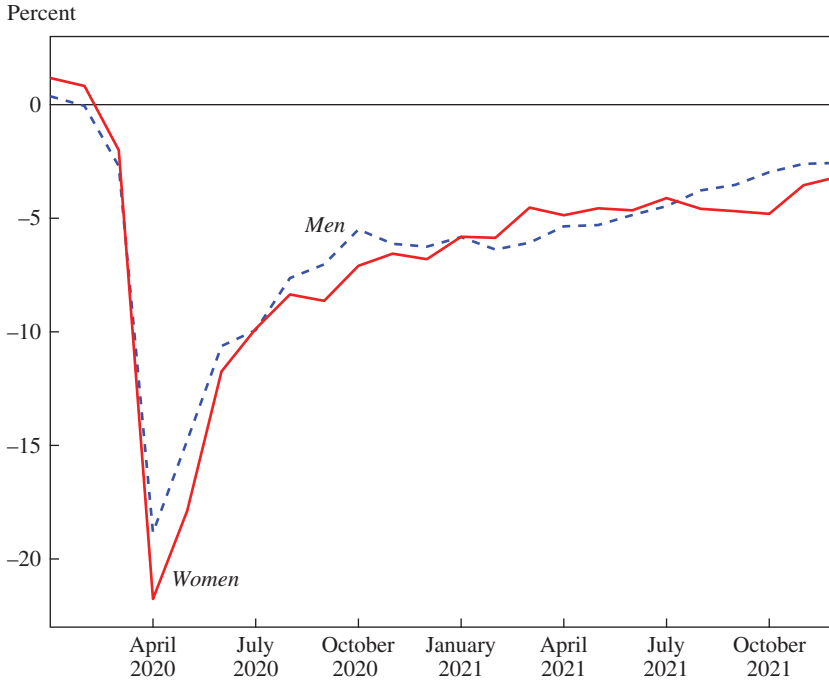
STEFANIA ALBANESI The COVID-19 pandemic has upended our lives and disrupted the economy and labor markets in many different ways. One reason it has been hard to grapple with the labor market impacts of the COVID-19 recession is its unique nature. Economic downturns in the United States are usually associated with a larger employment drop for men than for women, but during the COVID-19 recession, employment losses were larger for women (Albanesi and Kim 2021).

This is illustrated in figure 1 which reports the percentage change in the employment-to-population ratio by gender relative to the same month in 2019 for each month in 2020 and 2021. In April 2020, employment was 18 percent lower for men and 23 percent lower for women relative to April 2019. In October 2020, employment was 6 percent lower than in October 2019 for men and 8 percent lower for women. By December 2021, employment was still 3 percent lower relative to December 2019 for men and 4 percent lower for women.

There are demand-side and supply-side reasons for the gender differences in employment changes during typical recessions and during the COVID-19 recession. On the demand side, the asymmetry is partly explained by gender differences in the occupation distribution, with men primarily employed in production occupations and women concentrated in service occupations, which tend to be less cyclical (Albanesi and Şahin 2018). During the pandemic, however, there has been a sizable drop in the demand for services as a result of both the mitigation measures initially enacted to contain the pandemic and consumers' response to the risk of infection (Chetty and others 2020). Given the concentration of women in service

Brookings Papers on Economic Activity, Spring 2022: 111–139

Figure 1. Change in the Employment-to-Population Ratio Relative to the Same Month in 2019



Source: Author's calculations from the Bureau of Labor Statistics Current Population Survey.
Note: Population age 25–54 years old.

occupations, they have been disproportionately hit by the corresponding employment losses.

On the supply side, married women historically tend to increase their attachment to the labor force during economic downturns relative to expansions as a form of family-level insurance against the risk of employment loss for their husbands (Ellieroth 2019). This mechanism acts as an automatic stabilizer, and as the share of women in the labor force increased in the postwar period, it contributed to a reduction in the business cycle volatility of aggregate employment (Albanesi 2019). By contrast, during the pandemic, limited availability of in-person childcare and schooling options led some parents—and mothers in particular—to exit the labor force (Albanesiand Kim 2021).

Goldin challenges the notion that women experienced a disproportionate impact in the COVID-19 recession. She argues that differences by education were much larger than differences by gender and that some mothers may have been able to continue working due to the switch to working from

home, whereas they might have exited the labor force without the pandemic. In my discussion, I will qualify and contextualize these statements. I show that during the pandemic, employment did fall more for women, mostly because women were more likely to be working in jobs that could not be performed remotely. These in-person occupations mostly employ workers without a college degree, so college workers were mostly spared. I also find that Black and Hispanic workers are overrepresented in occupations that must be performed in-person, and this can account for racial disparities in employment outcomes. The distribution of workers across occupations does not fully account for the gender differences in employment losses, and labor supply also plays a role. I show that mothers did leave the labor force more than comparable fathers, but the rise in nonparticipation was mainly driven by workers who became unemployed first. So while the limited in-person childcare and schooling options likely contributed to the decline in labor force participation of mothers, those who were able to remain employed managed to continue working despite these challenges.

The decline in women's employment during the COVID-19 recession has raised concerns that the pandemic may lead to a long-lasting setback in women's employment going forward. To end my discussion, I will examine the potential for a jobless recovery as we exit the pandemic.

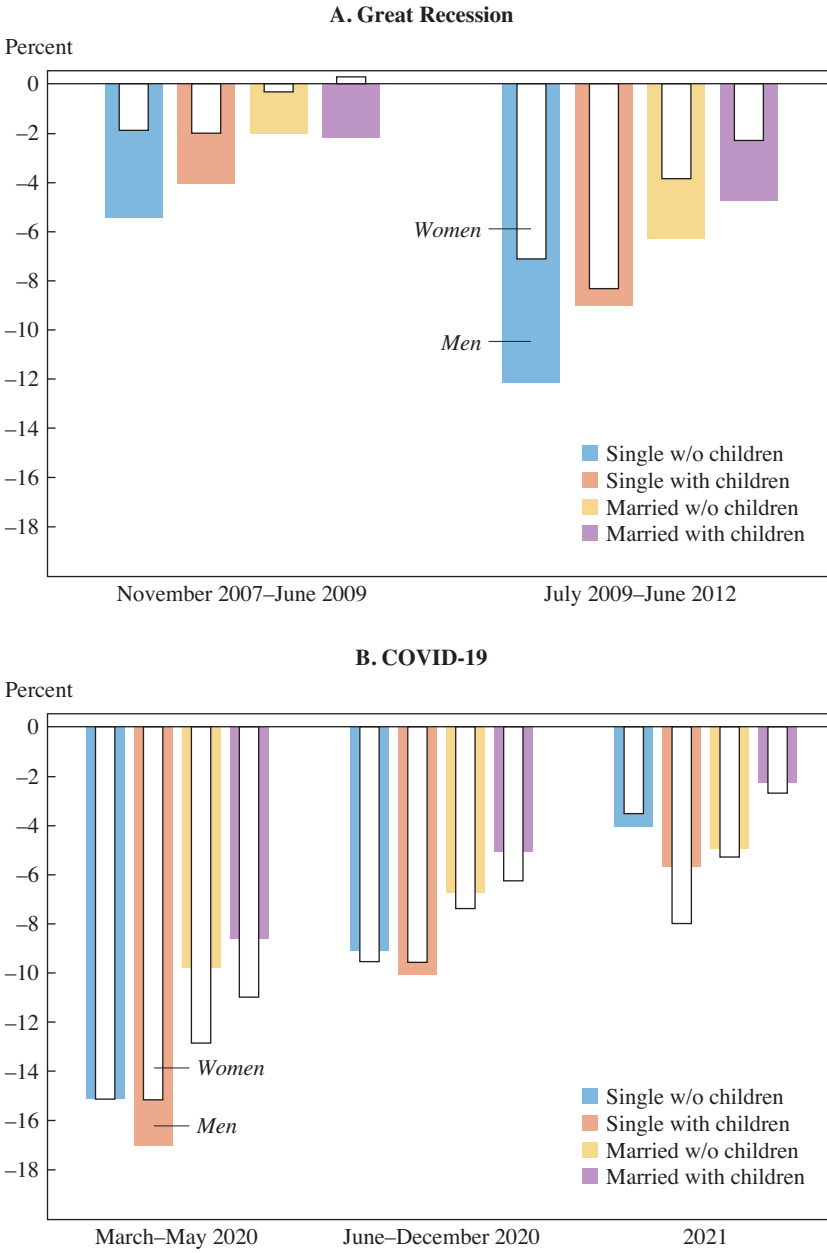
COMPARING COVID-19 TO THE GREAT RECESSION To gain perspective on how the employment losses of men and women during the COVID-19 recession differed from earlier recessions, it is useful to compare it to the Great Recession, which had a typical pattern.

Figure 2 shows the change in the employment-to-population ratio by gender and family status during COVID-19 and the Great Recession relative to prerecession values. For the Great Recession the comparison point is the same month in 2006, while for the pandemic recession it is the same month in 2019.[1] I divide the population into four demographic groups, by marital status and presence of children younger than 12 years old residing in the household.[2] For the Great Recession, I consider two phases. The first

1. These results are very similar if the same month in 2018 is used as a basis for COVID-19. Goldin argues that 2018 should be used as a counterfactual. However, when analyzing standard recessions, the most recent comparable prerecession date is used to measure the depth of the recession, and I follow this approach in my discussion consistent with the business cycle literature.

2. The size of each demographic group varies by gender. In February 2020, among women 17 percent are single without children, 6 percent are single with children, 15 percent are married without children, and 14 percent are married with children. Among men, 17 percent are single without children, 2 percent are single with children, 15 percent are married without children, and 15 percent are married with children.

Figure 2. Change in the Employment-to-Population Ratio for the Great Recession and for the COVID-19 Recession



Source: Author's calculations from the BLS Current Population Survey.

Note: For panel A, ratios are relative to 2006; for panel B, ratios are relative to 2019. Individuals with children are those who have children younger than 12 years old residing in the household.

is November 2007 to June 2009, which corresponds to the official recession dates determined by the Business Cycle Dating Committee of the National Bureau of Economic Research. The second runs from July 2009 to June 2012, when the broader economy was recovering but labor markets were still stagnant. For COVID-19, I consider three phases. The first comprises March, April, and May 2020, when the pandemic started and the strictest mitigation measures were in place. The second corresponds to the period from June to December 2020, with lower infection rates and less stringent mitigation measures, and the third phase is 2021.

During the Great Recession, the decline in women's employment was substantially smaller than men's for every demographic group. In the period from November 2007 to June 2009, the magnitude of the drop in employment for single women was less than half of the drop for single men. For married women, employment barely changed while it declined by 5 percent for married men. In the period from July 2009 to June 2012, gender gaps in employment loss were smaller but still largely favored women.

During COVID-19, the pattern is markedly different. Gender gaps in employment are negligible for single workers without children but are sizable for single parents and married workers. For married workers, the gender gaps were largest in March–May 2020, when married women experienced a decline in employment that was approximately 3 percent larger than for comparable men, and declined later in the pandemic. Among single parents, mothers experienced a slightly smaller decline in employment compared to single fathers in 2020, but during 2021 employment was 8 percent lower for single mothers and 6 percent lower for single fathers when compared to 2019.

Overall, the data support the notion that the decline in employment was larger for women during the pandemic, even if men also experienced substantial job loss. While gender gaps in employment losses were initially large but closed over time for married individuals, the pattern was reversed for single parents. The fact that in typical recessions women's job losses are smaller compared to men's and married women's employment typically does not decline likely triggered the alarm apparent in media and social commentary on the adverse effects of the pandemic on women's labor market performance.

LABOR DEMAND Labor demand and labor supply factors also contributed to women's larger employment losses during the pandemic. To explore the role of labor demand, Albanesi and Kim (2021) classify workers by occupation based on their flexibility and contact intensity. Flexible

Table 1. Occupation Classifications

	Flexible	*Inflexible*
High contact	Education, training, and library	Health care practitioners and technical health care support Food preparation and serving Personal care and service
Low contact	Management Business Computer and mathematical Architecture and engineering Life science, physical science, and social science Community and social services Legal Arts, design, entertainment, sports, and media Sales and related Office and administrative	Protective service Building and grounds cleaning and maintenance Farming, fishing, and forestry Construction trades, extraction Installation, maintenance, and repair Production Transportation and material moving

Source: Albanesi and Kim (2021); reproduced with permission by *Journal of Economic Perspectives*, © American Economic Association.

Note: Author's classification based on O*NET. Occupations are inflexible if they cannot be performed remotely, flexible otherwise. Occupations are high contact if they require interactions with coworkers or customers at a distance of less than six feet, low contact otherwise.

occupations include those that allow their employees to work remotely, whereas inflexible occupations require physical presence due to on-site equipment or outdoor activities. The distinction between high contact and low contact occupations is based on workers' physical proximity to customers or coworkers while on the job.

Table 1 displays where various occupations fall in the categorization, and table 2 reports the distribution of workers by gender across occupations prepandemic for the four categories defined in table 1. The inflexible, high contact occupations, comprising health, personal care, and hospitality, are the most vulnerable to lower demand due to COVID-19; they account for 17 percent of total employment and are dominated by female workers, with a women's share of 73 percent. Male workers are disproportionately represented in inflexible, low contact occupations, which account for 26 percent of total employment with a women's share of only 19 percent. Occupations in this category, comprising production, transportation, construction, and the like, experience the largest decline in employment in typical recessions. Flexible, low contact occupations are the largest category, accounting for 51 percent of overall employment, with women's share at 50 percent. These cover most professional and managerial jobs. Flexible, high contact

Table 2. Occupational Distribution by Gender

	Employed women	*Employed men*	*Total employed*	*Women's share*
Flexible, high contact	10	3	6	76
Flexible, low contact	53	48	51	50
Inflexible, high contact	26	9	17	73
Inflexible, low contact	11	40	26	19

Source: Albanesi and Kim (2021); reproduced with permission by *Journal of Economic Perspectives*, © American Economic Association.

Note: Values in percentages for February 2020.

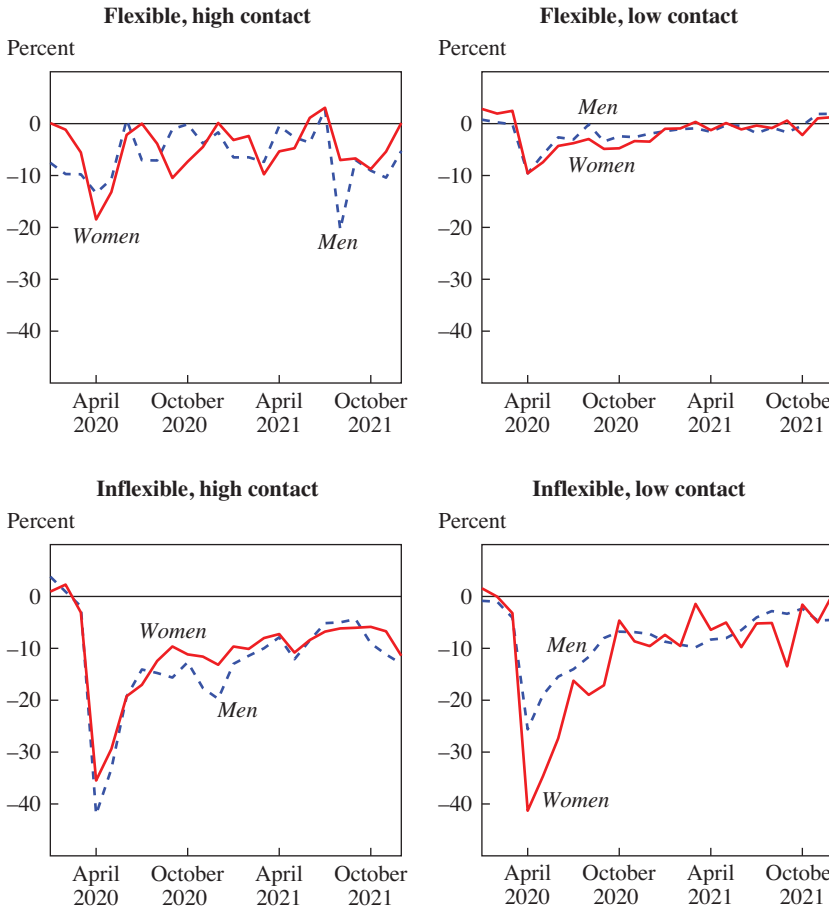
occupations, including most education jobs, also exhibit a high women's share at 76 percent but account for only 6 percent of total employment.[3]

Figure 3 displays the change in the employment-to-population ratio for these four occupational categories relative to the same month in 2019 by gender. Inflexible, high contact occupations show the largest decline in employment, with a drop in April 2020 relative to April 2019 of 38 percent for women and 41 percent for men, hovering at around −10 percent relative to the same month in 2019 from September 2020 until the end of 2021. Inflexible, low contact occupations are the second-worst hit, with a decline in employment close to 23 percent for men and 41 percent for women in April 2020 relative to April 2019. For these occupations, too, the recovery has stalled, with employment approximately 10 percent lower than in the same month in 2019 from October 2020 to the end of 2021. Employment in flexible, high contact occupations was 19 percent lower for women and 15 percent lower for men relative to one year prior in April 2020, but it recovered rapidly and has remained 2–8 percent lower than prepandemic levels from July 2020 onward. Finally, flexible, low contact occupations, which account for the biggest share of employment, were the least impacted, with a drop in employment of −10 percent relative to one year prior in April 2020 for both men and women, and a recovery to 2–4 percent lower relative to prepandemic levels from June 2020 onward.

Two patterns clearly emerge. First, for the flexible occupations, the decline in employment and the gender differences in that decline were small. The second pattern is that in inflexible occupations, initial employment losses were sizable and even at the end of 2021 employment remained well below 2019 values. Additionally, workers with the lowest representation

3. The occupation and industry distribution by gender does not vary by marital status; see Cortes and Pan (2018).

Figure 3. Percentage Change in the Employment-to-Population Ratio from Same Month in 2019 by Occupation

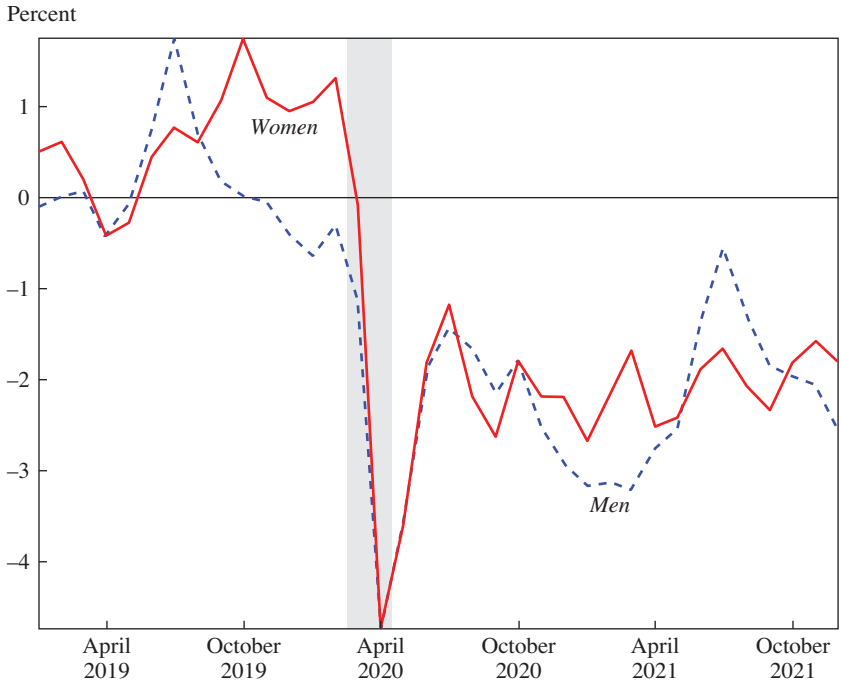


Source: Author's calculations based on the BLS Current Population Survey.

Note: Population age 25–54 years old. The numerator consists of the number of persons employed for each gender in each occupation, the denominator the number of persons of the same gender in the population.

by gender lost more jobs. This may be due to negative selection of male workers into female-dominated inflexible, high contact occupations and of female workers into the male-dominated inflexible, low contact occupations. The flexible occupations comprise most professional, managerial, and education jobs, and college-educated workers are disproportionally represented in these occupations, whereas inflexible occupations are dominated

Figure 4. Labor Force Participation Rate by Gender, January 2019–December 2021

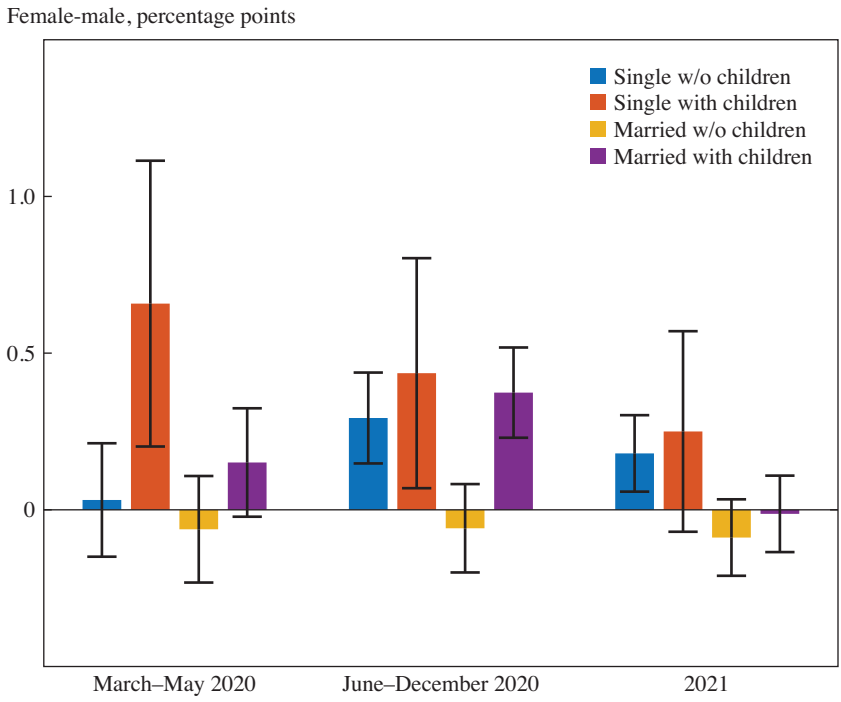


Source: Author's calculations based on the BLS Current Population Survey.
Note: Population age 25–54 years old. Percentage change since same month in 2018.

by workers without a college degree (Albanesi and Kim 2021). The difference in employment declines between flexible and inflexible occupations is much larger than the gender difference within occupations and matches the disparate effects by education highlighted in Goldin's work. But it is not the education per se that matters for the employment losses; rather college-educated workers were employed in jobs that can be performed remotely and that limited their employment losses.

LABOR SUPPLY Labor force participation declined for both men and women during the pandemic, as can be seen in figure 4, which plots the percentage change relative to the same month in 2018 in the labor force participation rate for the prime-age population in each month of 2019, 2020, and 2021. For both men and women, labor force participation was 5 percent lower in April 2020 compared to April 2018. Gender gaps in the change in participation relative to 2018 are on the order of zero to 2 percentage points during the pandemic and vary in sign. For both men and women, the participation

Figure 5. Female-Male Difference in Changes in Nonparticipation Relative to the 2019 Average by Family Status



Source: Author's calculations based on the BLS Current Population Survey.

Note: Controlled for age, education, and occupation. Error bars denote 90 percent confidence intervals. Population age 25–54 years old. Individuals with children are those who have children younger than 12 years old residing in the household.

rate had not recovered to prepandemic levels at the end of 2021, when it was still approximately 2 percent below the same period in 2018.

Both men and women experienced a surge in participation in late 2019 and the first two months of 2020 relative to three years prior, with a longer and more pronounced rise for women. Goldin correctly points out that using February 2020 as a basis for calculating the decline in participation overstates that decline for women. However, I will show that despite this caveat, nonparticipation rose more for women than for men.

Figure 5 presents female-male differences in the change in nonparticipation during the pandemic relative to the average in 2019 by family status, controlling for differences in age, education, and occupation across these groups. The estimates suggest that the biggest gender differences occur for

single parents in 2020 and married parents in the second half of 2020 and that by 2021 there are no longer sizable gender gaps.

Further breakdown of the data suggests that the rise in women's nonparticipation relative to men is mostly accounted for by transitions from unemployment rather than voluntary quits. This can be seen in figure 6, which reports the gender differences in the change in employment-to-nonparticipation and unemployment-to-nonparticipation flows, controlling for age, education, and occupation. There are sizable and significant gender differences for single parents in 2020 and for married parents in the second half of 2020 and in 2021. This is surprising, as it follows several decades of continued convergence in unemployment-to-nonparticipation flows across genders (Albanesi and Şahin 2018).

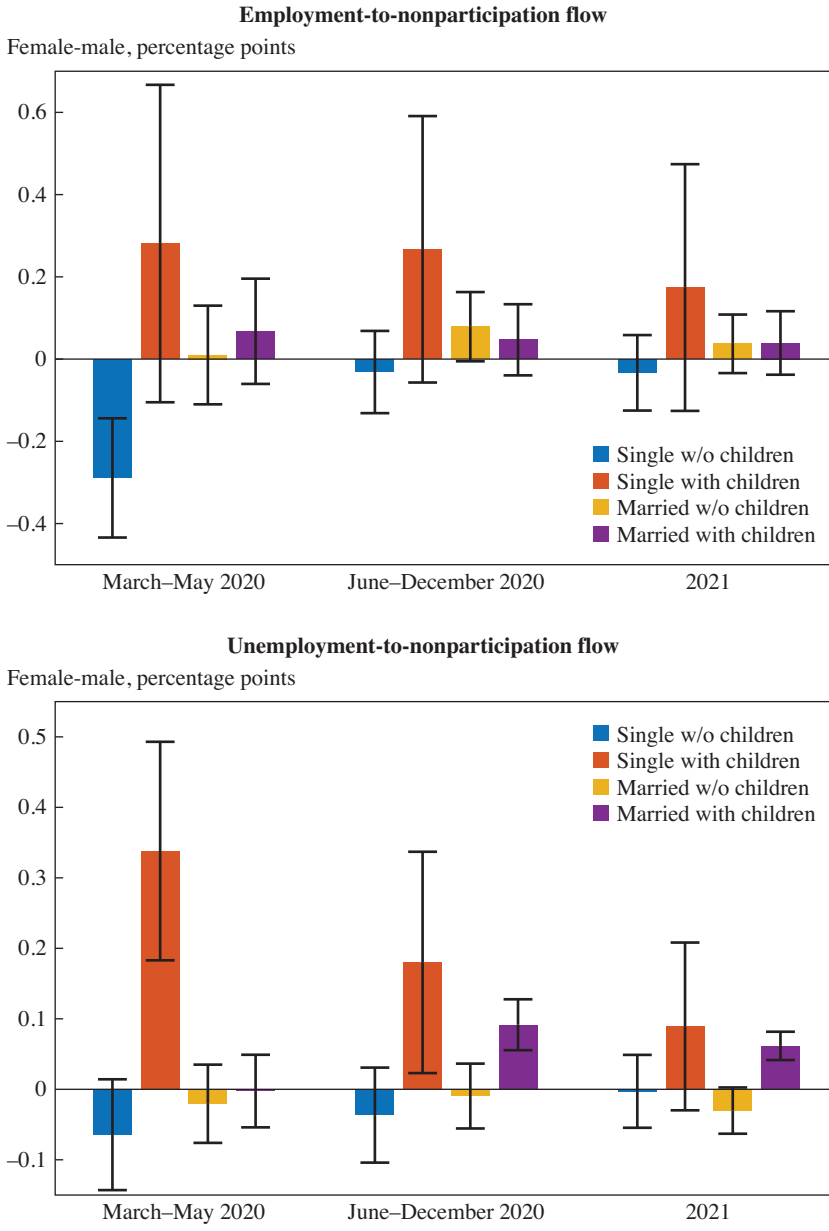
This finding suggests a pattern in which mothers who were able to keep their jobs during the pandemic continued working. However, those who lost their jobs exited the labor force at higher rates than comparable fathers.

RACIAL DISPARITIES The labor market impact of COVID-19 has been disparate by race. Figure 7 plots the change in the employment-to-population ratio relative to the same month in 2019 by race for men and women starting in January 2020, illustrating the large racial disparities for both men and women.

For men, at the start of the pandemic in spring 2020 the main difference is between white men, who experienced a 17 percent drop in employment, and the other racial groups, whose employment fell by 22–23 percent. During the rest of 2020, employment recovered more for Asian and white men, while during 2021 employment of Hispanic men converged to employment for white men but remained lower for Black men. Hispanic women experienced the most severe impact at the height of the pandemic, with a 28 percent decline in employment in April 2020 compared to the same month in 2019, while Black women experienced a 23 percent decline and white and Asian women a 20 percent decline. Asian, Black, and Hispanic women experienced a much slower recovery in employment during the rest of 2020, while in 2021 it was Black and Hispanic women's employment that lagged employment for both white and Asian women. Interestingly, gender gaps in the decline in employment are smallest for the Asian and Black population and largest for the white and Hispanic population.

What drives these racial disparities? My previous analysis suggests two possible economic factors. The first is the occupation distribution, which affects labor demand. Table 3 reports the occupation distribution for men and women by race. Focusing on women, we see that 28 percent of Asian women and 29 percent of Black women were employed in inflexible, high

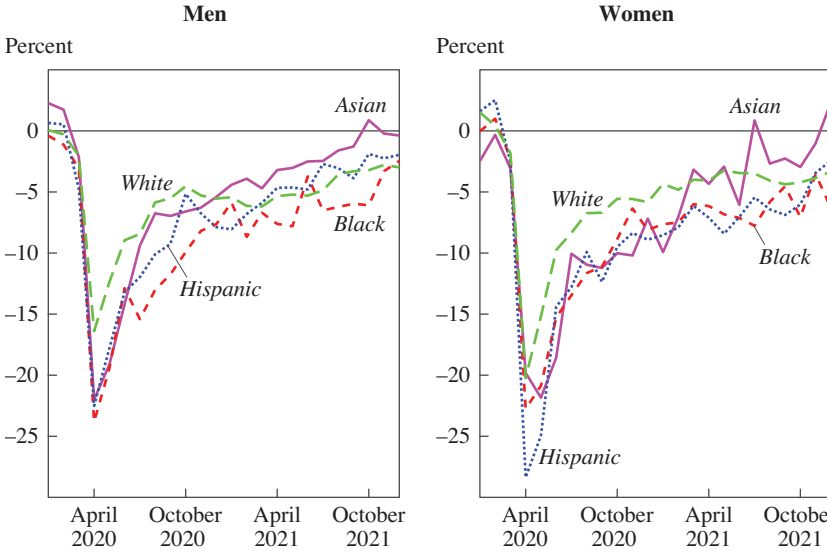
Figure 6. Female-Male Difference in Changes in Employment-to-Nonparticipation Flows and Unemployment-to-Nonparticipation Flows Relative to the 2019 Average by Family Status



Source: Author's calculations based on the BLS Current Population Survey.

Note: Controlled for age, education, and occupation. Error bars denote 90 percent confidence intervals. Population age 25–54 years old. Individuals with children are those who have children younger than 12 years old residing in the household.

Figure 7. Change in the Employment-to-Population Ratio Relative to the Same Month in 2019 by Gender and Race, January 2020–December 2021



Source: Author's calculations based on the BLS Current Population Survey.

Table 3. Occupation Distribution by Race and Gender

	Asian		*Black*		*Hispanic*		*White*	
	Men	*Women*	*Men*	*Women*	*Men*	*Women*	*Men*	*Women*
Flexible, high contact	3	7	3	7	2	7	4	12
Flexible, low contact	65	58	40	50	31	46	53	56
Inflexible, high contact	14	28	11	29	9	24	7	24
Inflexible, low contact	18	7	46	14	58	23	36	8

Source: Author's calculations based on the BLS Current Population Survey.
Note: Percentage in each occupation by gender/race in February 2020.

contact occupations, compared to 24 percent of Hispanic and white women. Additionally, 23 percent of Hispanic women were employed in inflexible, low contact occupations, compared to 7 percent of Asian women, 14 percent of Black women, and 8 percent of white women. By contrast 58 percent of Asian women and 56 percent of white women were employed in flexible, low contact occupations compared to 50 percent of Black women and 46 percent of Hispanic women. This suggests that the overrepresentation of Black and Hispanic women in inflexible occupations and the overrepresentation of Asian and white women in flexible occupations contributed to racial disparities in employment.

The second possible factor is family status, which affects labor supply. As previously noted, the rise in nonparticipation during the pandemic was most pronounced for single mothers, followed by married mothers. Twenty-two percent of Black women are single mothers, compared to 16 percent for Hispanic women, 8 percent for white women, and 4 percent for Asian women. Additionally, the fraction without children, combining both single and married, is the lowest for Hispanic women, at 55 percent, while it is above 60 percent for the other racial groups.[4] The higher incidence of single mothers among Black women and of both single and married mothers among Hispanic women may have contributed to a bigger reduction in their labor supply, compared to white and Asian women during COVID-19.

Another important factor affecting labor supply is the incidence of COVID-19 infections, particularly severe cases requiring hospitalization, across racial groups. It is well documented that COVID-19 infection rates have been higher in Black and Hispanic communities, throughout the course of the pandemic. Goldin's work shows that this depressed the labor supply of women in these groups, both directly due to their exposure to disease and indirectly, through the rise in care needs. The impact of COVID-19 infections and the resulting sequelae on labor supply has not been addressed in economic research and is an important topic for future work.

WILL THE JOBS RETURN? Figures 1 and 4 clearly show that at the end of 2021 both employment and participation had not regained prepandemic values for both men and women. As we look forward to the end of the pandemic, one critical question is whether the labor market will fully recover. Since the 1990–1991 recession, the United States has experienced jobless recoveries, that is, even as GDP and aggregate demand rebounded, labor markets continued to stagnate and employment struggled to attain prerecession levels.

There are two main explanations for jobless recoveries. The first is that the slow and incomplete rebound of employment was due to the adoption of labor-saving technologies, such as automation, leading to a long-run decline in the demand for routine jobs. The resulting job losses are concentrated in recessions, and when the economy recovers, the lost jobs are not reinstated. This phenomenon affects primarily middle-skill workers and is a key mechanism through which the trend toward job polarization has affected business cycles (Acemoglu and Autor 2011; Jaimovich and Siu 2020).

The effects of the pandemic have been mostly felt in service occupations that may seem less amenable to automation. However, the pandemic

4. The distribution of family status of women by race was calculated from the Current Population Survey.

has also given employers an additional incentive to embrace automation, as long as the risk of COVID-19 infection persists. Are jobs that were lost during the COVID-19 recession more or less susceptible to automation?

One way to measure the susceptibility to automation is routine task intensity (RTI), an index developed by Autor and Dorn (2013), which calculates the task inputs in each occupation based on job requirements. Albanesi and Kim (2021) calculate that 34 percent of all jobs in inflexible, high contact occupations were highly susceptible to automation in February 2020, compared to 22 percent of all jobs in inflexible, low contact occupations that are most hit by typical recessions. These findings raise the possibility that employment losses in those occupations may not be fully reversed as the broader economy recovers from the pandemic.

The second explanation for jobless recoveries is the flattening of female labor force participation starting in the early 1990s. In my own work, I have shown that, even before the 1990s, recoveries had been jobless for men. However, as long as female labor force participation was rising briskly, women's employment tended to grow very rapidly in recoveries, sustaining aggregate employment (Albanesi 2019). As the rise in female participation slowed in the 1990s, the growth of women's employment during recoveries became similar to men's, slowing the recovery of aggregate employment. Not only that, but the rise in women's participation, while it lasted, sustained productivity and GDP growth and contributed to increased men's wages. Given the critical role of women's participation for aggregate economic performance in the United States, macroeconomists and policymakers should track this indicator closely and seek to understand its behavior in the trend and over the business cycle.

REFERENCES FOR THE ALBANESI COMMENT

Acemoglu, Daron, and David H. Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics, Volume 4B*, edited by David Card and Orley Ashenfelter. Amsterdam: North Holland.

Albanesi, Stefania. 2019. "Changing Business Cycles: The Role of Women's Employment." Working Paper 25655. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25655.

Albanesi, Stefania, and Jiyeon Kim. 2021. "Effects of the COVID-19 Recession on the US Labor Market: Occupation, Family, and Gender." *Journal of Economic Perspectives* 35, no. 3: 3–24.

Albanesi, Stefania, and Ayşegül Şahin. 2018. "The Gender Unemployment Gap." *Review of Economic Dynamics* 30:47–67.

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review*103, no. 5: 1553–97.

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.

Cortes, Patricia, and Jessica Pan. 2018. "Occupation and Gender." In *The Oxford Handbook of Women and the Economy*, edited by Susan L. Averett, Laura M. Arguys, and Saul D. Hoffman. Oxford: Oxford University Press.

Ellieroth, Kathrin. 2019. "Spousal Insurance, Precautionary Labor Supply, and the Business Cycle—A Quantitative Analysis." Meeting Paper 1134. Society for Economic Dynamics.

Jaimovich, Nir, and Henry E. Siu. 2020. "Job Polarization and Jobless Recoveries." *Review of Economics and Statistics* 102, no. 1: 129–47.

COMMENT BY

JANE OLMSTEAD-RUMSEY In this paper, Goldin presents a careful and wide-ranging analysis of women's employment experiences during the coronavirus pandemic, assessing and in some cases correcting popular narratives that have developed about the pandemic's effects on women in the labor market. She examines the role that education levels, childcare responsibilities, telecommuting, occupations, and race have played in shaping labor market outcomes during the pandemic. She argues that the decline in the female labor force participation rate during the pandemic was not large relative to the historical average since the late 1980s, when the increase in women's labor force participation began to slow, and that estimates of the decline depend significantly on the reference month chosen. A robust finding is the rise in caregiving time by women during the pandemic.

My discussion concerns three primary issues. The first of these considers the appropriate counterfactual for labor force participation rates absent the pandemic, including a closer examination of the rise in female labor force participation prior to the pandemic. The second relates to the long-run impact of the pandemic on women through changes in the availability of remote work. The third considers policy implications of "she-cessions" compared to "man-cessions."

LABOR FORCE PARTICIPATION RATES Prior to the start of the pandemic, the US economy had been in a long expansionary period. In August 2020 the Federal Reserve announced changes to its long-run monetary policy strategy, explicitly describing its maximum employment mandate as a "broad-based

and inclusive" goal.[1] In a speech accompanying this announcement Chairman Powell (2020) noted that marginalized groups were benefiting more from the booming labor market as time went on. It is impossible to know how long the expansion would have continued absent the pandemic, but it is worth thinking hard about the right counterfactuals for labor force participation rates when considering how large the impact of the pandemic was on different groups of women compared to a world with no pandemic. My first comment is therefore an investigation of the rise in women's labor force participation prior to the pandemic.

The paper shows that the labor force participation rate (LFPR) for women age 25–54 increased rapidly in late 2019. This surge makes the choice of a reference month for measuring declines in female labor force participation during the pandemic important. The paper rightly notes that women's LFPR did not grow significantly in the past three decades.[2] However, since 1988, there have been significant fluctuations between 72.2 percent and 77.3 percent, corresponding with the business cycle, so understanding where women's LFPR was heading before the pandemic is still important.

I note five features of the rise in women's LFPR prior to the pandemic which suggest it would likely have remained high absent the pandemic: First, it was sustained. The largest monthly increase in women's LFPR in 2019 was 0.9 percentage points between July and August, from 75.4 percent to 76.3 percent. Rather than declining thereafter, which would reflect a possible statistical or seasonal anomaly, the rate continued to rise slowly through February 2020 to 76.9 percent. Second, it was part of a recovery in women's LFPR that began in late 2015, so a comparison to earlier years like 2018 makes losses due to the pandemic seem smaller. This is particularly true for subgroups that tend to enter later in the cycle. Third, it was driven by employment, not unemployment, rising—the unemployment rate for women was flat at around 3.5 percent in the later part of 2019 and early 2020.[3] Fourth, it was fairly broad-based among women. Following the subgroup analysis in the paper comparing changes in LFPR for subgroups in the Current Population Survey (CPS) between April 2019 and December

1. "Federal Open Market Committee Announces Approval of Updates to Its Statement on Longer-Run Goals and Monetary Policy Strategy," press release, Board of Governors of the Federal Reserve, https://www.federalreserve.gov/newsevents/pressreleases/monetary20200827a.htm.

2. One explanation for this is the rise in the college wage premium; see Albanesi and Prados (2022). Women's pay and labor force attachment over the life cycle have continued to increase since the 1980s, according to Goldin (2006) and Goldin and Mitchell (2017).

3. See FRED, "Unemployment Rate—Women," https://fred.stlouisfed.org/series/LNS14000002.

Table 1. Reduced Form Estimates for Labor Force Exit, March to the following April

	Labor force exit
College or more	0.029***
	[0.004]
Age 30–39	0.030***
	[0.005]
Age 40–49	0.038***
	[0.005]
Age 50–54	0.034***
	[0.006]
Non-white	0.023***
	[0.004]
Children, under age 5	0.024***
	[0.005]
Children, age 5–14	0.002
	[0.004]
Married	0.010**
	[0.004]
Group 1	0.076***
	[0.012]
Pandemic	0.027***
	[0.003]
Group 1 × pandemic	–0.024
	[0.018]
Constant	0.162***
	[0.034]
Industry × Occupation Fixed Effects	Yes
Number of observations	48,432
R^2	0.223

Source: Current Population Survey.

Note: Population is women age 25–54 only. Sample includes group 1 (new entrants in 2019), group 2 (continuously employed workers in 2019), and placebo group 1 and placebo group 2 for placebo pandemic in 2018 (see note 21 in the paper). Weighted using "wtfinl" from the CPS micro data. Robust standard errors in brackets.

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

2019 by age, education, and presence of children under age 5, I find that all subgroups except three saw increases in LFPR of at least 0.9 percentage points.[4] Fifth, as of January 2022, women's LFPR had recovered to 76 percent, a higher level than any month between February 2009 and July 2019, suggesting that trend LFPR was high.

Turning to micro data, I follow the creation of group 1 and group 2 in the pandemic and placebo periods using CPS data as in the paper. As the paper points out (table 1), these groups differ substantially on observables.

4. The three groups without large increases were noncollege-educated women in their twenties without children under age 5 (increase of 0.3 percentage points), college-educated women in their forties without children under age 5 (decrease of 0.7 percentage points), and noncollege-educated women in their fifties without children under age 5 (decrease of 0.1 percentage points).

Women who entered the labor force just prior to the pandemic (group 1) are younger, more likely to have children, and less likely to have a college degree. These were precisely the groups at greatest risk of leaving their jobs during the pandemic because of lack of remote work opportunities and because of childcare needs. I therefore run a regression to control for demographic differences between the two groups:

$$y_i = \alpha + \beta X_i + \delta 1^i \text{ (group 1, group 1 placebo)} + \gamma 1^i \text{ (pandemic groups)} + \rho 1^i \text{ (group 1)} + \epsilon_i,$$

where y_i is an indicator for whether woman i left the labor force any time between March and the following April, X_i is a vector of demographic characteristics including age, race, presence of children, marital status, and occupation by industry dummies, and 1^i is an indicator for woman i's membership in a given group.

As shown in table 1, controlling for demographics and job types shrinks the difference in the propensity to leave the workforce between group 1 (new entrants in 2017 and 2019) and group 2 (continuously employed workers in 2017 and 2019), measured by δ, to 7.6 percentage points, still a large number reflecting the weaker attachment of new entrants to the labor force. Both group 1 and group 2 were more likely to leave the workforce during the pandemic than their counterparts in the placebo groups were to leave the workforce in March 2018–April 2019 (the estimated value of γ is 2.7 percentage points), but the interaction between being in group 1 and the pandemic is not significant, suggesting that women who entered the workforce just before the pandemic were not more likely to leave the workforce during the pandemic than similar women were to leave the workforce in March 2018–April 2019 after accounting for the effects of the pandemic itself.

Taken together, this analysis suggests that using the February 2020 women's LFPR as a reference month for measuring pandemic-related declines, as many papers have done, is not unreasonable.[5] Policymakers have recently devoted increased attention to the behavior of the LFPRs of marginalized groups over the business cycle. More work should be done to understand the long-run drivers of both male and female LFPRs to develop estimates of these trends going forward.

5. When looking at CPS subgroups, however, participation rates still need to be seasonally adjusted. This is not an issue for the headline series that is seasonally adjusted by the Bureau of Labor Statistics.

CHANGES IN WORK FLEXIBILITY The paper provides an insightful discussion of the potential benefits and risks for women of greater work flexibility after the pandemic. In Alon and others (2020b) we use a model of household labor supply featuring heterogeneity by gender, marital status, age of children, occupation (telecommuting or not), and human capital to understand how a permanently higher share of telecommuting jobs in the United States would have an impact on female labor force participation and the gender wage gap in the long run.

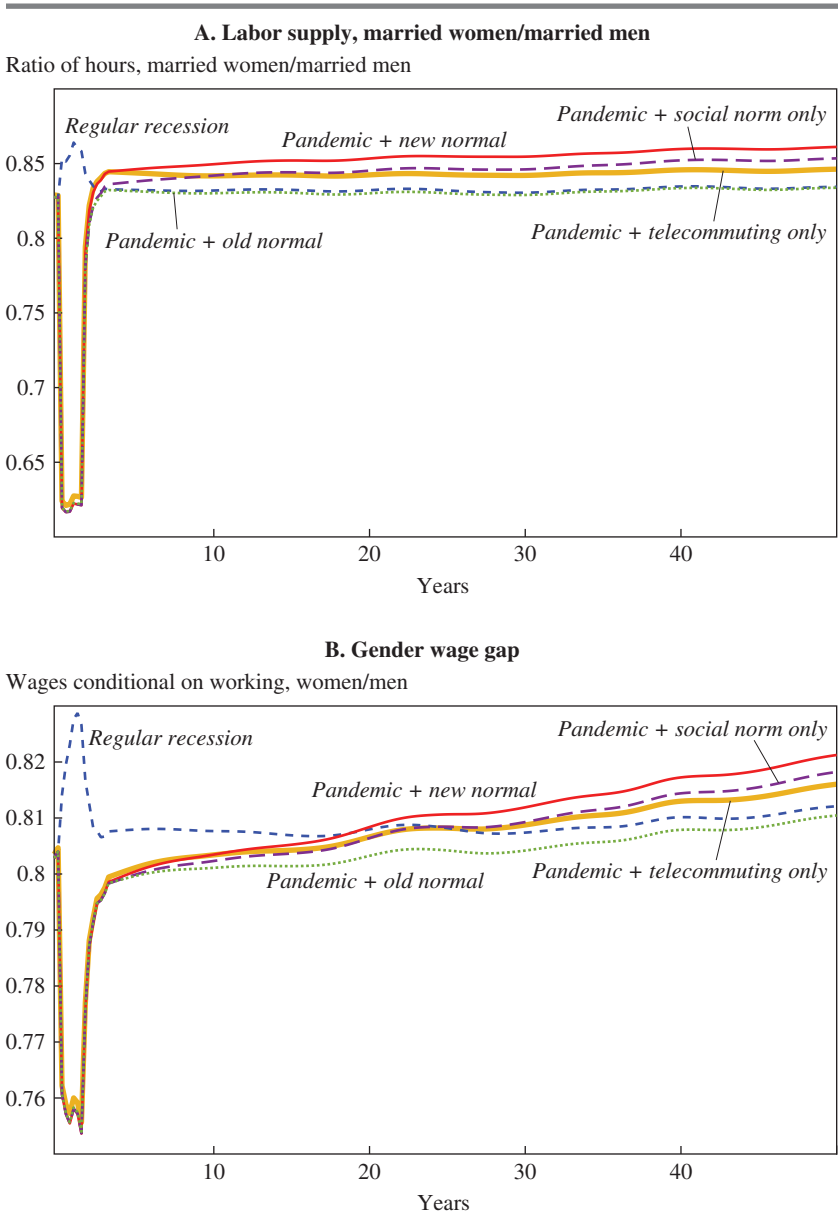
Crucially, we assume there are no productivity losses or gains associated with telecommuting compared to non-telecommuting jobs, only that telecommuting allows workers to combine a fraction of childcare time with work.[6] This fraction is calibrated to match the fact that fathers who could telecommute did 50 percent more childcare than fathers who could not, based on our estimates from precrisis time use data (Alon and others 2020a). Individuals who are not working risk human capital depreciation whereas those who are working full time accumulate human capital. We study a permanent rise in the fraction of the workforce in telecommuting occupations from 13 percent prepandemic to 30 percent in the so-called new normal after the pandemic ends.

Figure 1, panel A, shows the model prediction for the evolution of relative hours for married women compared to married men as a result of this change ("Pandemic + telecommuting only"), suggesting that married women's relative hours could rise by about 2 percent in the long run if work flexibility persists. This also reduces the gender wage gap significantly in the long run by increasing women's labor force attachment and accumulated human capital (figure 1, panel B). "Pandemic + old normal" shows the counterfactual where telecommuting reverts to about 13 percent of jobs and shows that this would significantly slow the recovery of the gender wage gap to prepandemic levels.

The latest data from Barrero, Bloom, and Davis (2021) show that men and women who can work from home are currently working the same number of days per week at home. The difference between desired work from home days per week for men and women is small, with women hoping for 2.37 and men for 2.18, but men report higher work from home days allowed by their employer (1.42 versus 1.24 for women). In an analysis of

6. The evidence on this is mixed. As discussed in the paper, female academics have been less productive during the pandemic. Women also report more interruptions while working from home (Andrew and others 2021). But workers overall self-report higher productivity while working from home (Barrero, Bloom, and Davis 2021).

Figure 1. Married Women's Relative Hours and the Gender Wage Gap Conditional on Working



Source: Alon and others (2020b); reproduced with permission.

Note: "Pandemic + telecommuting only" shows effects of a pandemic recession (equal job loss risk for men and women and a large increase in childcare time) with a permanent increase in the fraction of telecommuting jobs to 30 percent. "Pandemic + old normal" shows the case where the fraction reverts to 13 percent after the pandemic.

prepandemic time use data we found that while married men were more likely than married women to be able to telecommute (45 percent versus 42 percent), married men telecommuted 25 percent less than married women (thirty days per year for married men versus forty-one days per year for married women; Alon and others 2020a). It remains to be seen whether the pandemic has changed norms around this behavior. If not, the risks to women's face time and promotions remain.

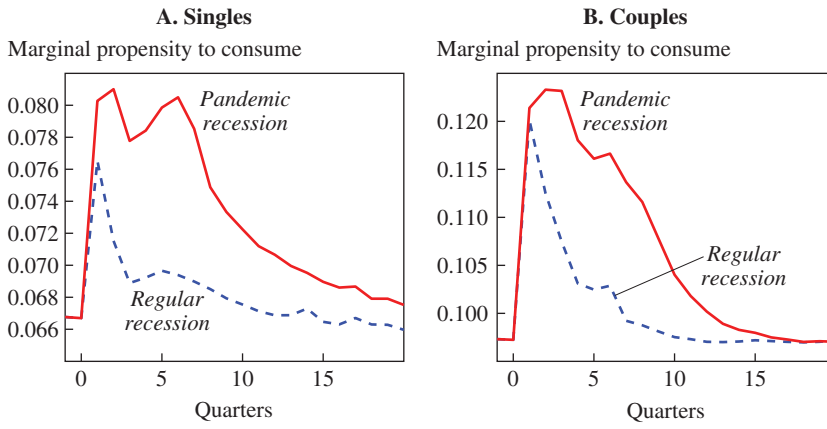
The media headlines at the time of writing regarding women's employment during the pandemic relate to the "Great Resignation." In many cases these stories suggest that the recent rise in quitting is due to workers, especially women, seeking greater employment flexibility. Data on quitting and job-to-job transitions by gender for 2021 are not yet available, nor will it be clear from these data how many workers quit in pursuit of jobs with greater flexibility, but research and surveys on this issue will measure another important dimension of women's labor market experiences during the pandemic related to reallocation to more flexible jobs.

POLICY IMPLICATIONS The 2020 recession due to COVID-19 reversed a pattern from the last five recessions in terms of the gender composition of job losses (Alon and others 2020b), creating distinct policy considerations for she-cessions versus man-cessions. First, she-cessions likely feature a greater translation of employment losses to consumption because of a decrease in intrafamily insurance.[7] In typical recessions, which have been man-cessions, married women can increase their labor supply if their husband loses his job. Because female-dominated occupations were most affected by the 2020 recession, and because of childcare needs, this option was less available than in previous recessions, and most men were already working full time and could not further increase hours in response to their wife's job loss. However, because of the fiscal response to the pandemic, many households that experienced drops in labor income also experienced offsetting increases in government transfers from unemployment insurance and stimulus checks, making measurement of this issue in the data difficult.

Using our model (Alon and others 2020b), we find that households' marginal propensities to consume (MPCs) are particularly high in pandemic recessions compared to regular recessions because the family insurance channel is diminished (figure 2). All else equal, elevated MPCs imply greater efficacy of fiscal stimulus. The paper points out that differences in job loss rates between education levels are even starker than gender gaps,

7. Bardóczy (2022) finds that this family insurance channel reduces the volatility of aggregate consumption in the United States by 33.5 percent.

Figure 2. Average Marginal Propensities to Consume



Source: Alon and others (2020b); reproduced with permission.
Note: The average fraction of an unexpected transfer that a household would consume instead of save.

and similar considerations apply when thinking about recessions that disproportionately affect people with less education, who typically have less savings to insure themselves against negative income shocks. Thus, amplification may be larger in these sorts of recessions as well, and the effects of fiscal policy greater.

A second policy consideration is "scarring" of workers who lost their jobs during the recession in terms of future employment and earnings. Lifetime earnings losses from job displacement are especially large for people who lose their jobs during recessions (Davis and von Wachter 2011). Furthermore, Alon and others (2022) note that at the micro level women's labor supply is more elastic than men's, so women who have dropped out due to the pandemic will also likely take longer to reenter the workforce or never return, making the recovery of employment even slower than it has been after previous recessions. Of course, the recovery of women's LFPR noted earlier looks strong compared to 2019, but it may be well below the prepandemic trend. In other countries, like Germany and the United Kingdom, employment relationships between workers and firms were preserved to a greater extent than in the United States because of short-time work policies (Alon and others 2022), which is one possible way to mitigate scarring in future recessions.

Finally, it's worth noting that comparing a man-cession and a she-cession with equal numbers of job losses for men in the man-cession and women in

the she-cession, we would expect the man-cession to feature a larger drop in aggregate labor income because men are paid more and work more hours on average than women. How exactly this translates into the decline in total output depends on the distribution of household MPCs, and the employment recovery from a she-cession will likely be slower. These considerations matter not just for pandemic recessions, which may also affect caregiving responsibilities, but for any recession where job losses are greater for women than for men, such as a recession concentrated in women-dominated service industries.

CONCLUSION The paper comprehensively characterizes the labor market experiences of women during the COVID-19 crisis. A central message is that despite fears that we would see a mass exodus of women from the workforce, most women remained in their jobs despite increased caregiving responsibilities for many. Education levels and race were important factors in determining which women left the workforce. Exactly how large the job losses and hours reductions were for women depends on the precrisis reference month chosen. I provide some reasons to believe the high LFPR of women prior to the pandemic would have been sustained if the crisis had not occurred, meaning that job losses were indeed substantial. Regardless of the exact percentage decline, the pandemic recession was distinct from previous recessions in terms of its larger effects on women than on men. This has implications for macroeconomic stabilization policies and the speed of the employment recovery back to trend. Models can help us forecast the long-run effects of employers continuing to allow employees to work from home, but these estimates depend on take-up of this flexibility by gender and on productivity and promotions at home versus in the office.

REFERENCES FOR THE OLMSTEAD-RUMSEY COMMENT

Albanesi, Stefania, and María José Prados. 2022. "Slowing Women's Labor Force Participation: The Role of Income Inequality." Working Paper 29675. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29675.

Alon, Titan, Sena Coskun, Matthias Doepke, David Koll, and Michèle Tertilt. 2022. "From Mancession to Shecession: Women's Employment in Regular and Pandemic Recessions." *NBER Macroeconomics Annual* 36:83–151.

Alon, Titan, Matthias Doepke, Jane Olmstead-Rumsey, and Michèle Tertilt. 2020a. "The Impact of COVID-19 on Gender Equality." Working Paper 26947. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26947.

Alon, Titan, Matthias Doepke, Jane Olmstead-Rumsey, and Michèle Tertilt. 2020b. "This Time It's Different: The Role of Women's Employment in a Pandemic Recession." Working Paper 27660. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27660.

Andrew, Alison, Sarah Cattan, Monica Costa Dias, Christine Farquharson, Lucy Kraftman, Sonya Krutikova, Angus Phimister, and Almudena Sevilla. 2021. "The Gendered Division of Paid and Domestic Work under Lockdown." Working Paper W21/17. London: Institute for Fiscal Studies. https://ifs.org.uk/publications/15497.

Bardóczy, Bence. 2022. "Spousal Insurance and the Amplification of Business Cycles." Working Paper. https://www.bencebardoczy.com/publication/bardoczy_jmp/bardoczy_jmp.pdf.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis. 2021. "Why Working from Home Will Stick." Working Paper 28731. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28731.

Davis, Steven J., and Till von Wachter. 2011. "Recessions and the Costs of Job Loss." *Brookings Papers on Economic Activity*, Fall, 1–72.

Goldin, Claudia. 2006. "The Quiet Revolution That Transformed Women's Employment, Education, and Family." *American Economic Review* 96, no. 2: 1–21.

Goldin, Claudia, and Joshua Mitchell. 2017. "The New Life Cycle of Women's Employment: Disappearing Humps, Sagging Middles, Expanding Tops." *Journal of Economic Perspectives* 31, no. 1: 161–82.

Powell, Jerome H. 2020. "New Economic Challenges and the Fed's Monetary Policy Review." Speech given at Navigating the Decade Ahead: Implications for Monetary Policy, Federal Reserve Bank of Kansas City, Jackson Hole, Wyo., August 27. https://www.federalreserve.gov/newsevents/speech/powell20200827a.htm.

GENERAL DISCUSSION Robert Hall noted that the people who accounted for the huge reduction of work in April 2020 did not all lose their jobs. He pointed out that the snap back from the pandemic recession was vastly faster than any other recession because workers were recalled to existing jobs. He explained that until it dissipated around early fall 2021, all other dynamics were dominated by the temporary layoffs. He stated that focusing on participation is appropriate because it includes unemployment, so even when people temporarily lost their jobs they were counted as part of the labor force.

Betsey Stevenson noted that people may have made accommodations and reduced work that does not show up as employment in the data. This is related to Claudia Goldin's points about childcare and to what Stevenson has seen in her own survey work, as well, that workers may be turning down

promotions, saying no to training, or putting in slightly fewer hours.[1] She explained that the resulting long-term effects on the career paths and earnings of these individuals cannot be seen yet. Stevenson noted that, surprisingly, her data show that men and women were equally affected, reflecting the fact that men were not satisfied with the time they got to spend with their children going into the pandemic and were able to change that to some extent. This may represent a shock to preferences and, related to what Jane Olmstead-Rumsey said in her discussion, opens up the question of whether men will take advantage of opportunities to work from home more—something which would be an important mark for gender equality.

Steven Davis explained that, in addition to the increase in unemployment resulting from temporary layoffs, there was a sizable atypical exit from the labor force during the pandemic. Those people were not on temporary layoff and will have a much longer path back to the labor force. Commenting on the concern in much of the discussion about what higher rates of working from home might mean for women in their career prospects, Davis mentioned his own work and a survey of working arrangements and attitudes.[2] He said that the data showed that many would prefer to work from home part of the week, and that when asked how much they value the option to work from home two or three days a week compared to an equivalent pay increase, the mean ranged from about 5 to 8 percent of earnings—a big number, he noted. He also found that 43 percent of respondents said they would look for another job that offers the ability to work remotely if their current employer mandated a five-day, in-person workweek.[3] Davis suggested that despite misgivings or concerns many—maybe most—people have a strong desire to work from home or to have the flexibility to do so part of the time. In response to Stevenson's point, Davis considered the gap between what people want and what they will get. Referring again to the survey, Davis and coauthors found that men, as part of their long-term arrangements at work, will be doing more remote work than women. So in terms of what workers want and what they will get when it comes to remote work, the gap

1. Betsey Stevenson, *Women, Work, and Families: Recovering from the Pandemic-Induced Recession* (Washington: The Hamilton Project, Brookings Institution, 2021), https://www.hamiltonproject.org/assets/files/COVID_Recovery_Stevenson_v5.pdf.

2. Jose Maria Barrero, Nicholas Bloom, and Steven Davis, "Why Working from Home Will Stick," working paper 28731 (Cambridge, Mass.: National Bureau of Economic Research, 2021), https://www.nber.org/papers/w28731; and the Survey of Working Arrangements and Attitudes, https://wfhresearch.com/.

3. Jose Maria Barrero, Nicholas Bloom, and Steven Davis, "Don't Force People to Come Back to the Office Full Time," *Harvard Business Review*, August 24, 2021.

is greater for women than for men. He closed by encouraging more thinking about the upside of the flexibility that comes with working from home.

Olmstead-Rumsey noted that, just as Davis mentioned, surveys have shown that workers prefer jobs with more flexibility and may even quit their current job in pursuit of it. At the same time, she noted, there is sometimes a difference between what people say and what they do. For example, some of the headlines that Goldin cited about women expressing their intention to scale back or not come back to the workforce at all have not come to fruition. She brought up the "Great Resignation" and then wondered whether the desire for flexible work arrangements had anything to do with it but noted the lack of data in this regard. In sum, she said, we should be mindful of the fact that while people express dissatisfaction in these surveys, their behavior may look quite different.

Stevenson considered what the right counterfactual might be and pointed to the steep job growth between 2015 and 2019 in the service sector, in the types of jobs that women tend to hold. Stevenson noted that in February 2020 there were almost a million fewer jobs in the goods-producing sector compared to prior to the Great Recession. However, Stevenson was worried that we hadn't seen a full recovery in demand for services yet.[4] She pondered to what extent we would be able to receive services similar to before the recession if workers did not return. She emphasized that this hinges on the choice of women to come back. She returned to her question about the right counterfactual and wondered whether looking at seasonally adjusted data for the end of 2019 or looking at growth rate projections between 2017 and 2019 was preferable.

Stefania Albanesi responded to Stevenson's remarks relative to automation across different industries and noted that in past recessions manufacturing may have been more susceptible to the adoption of labor-saving technologies than services, notably because of care services. Her own research on the susceptibility of automation across industries that were hardest hit by the pandemic recession, however, shows that about 33 percent of workers, mainly in service occupations, were highly susceptible to automation.[5] Albanesi explained that this number is very similar to that of manufacturing and construction

4. According to the Bureau of Labor Statistics, there were 14.5 million jobs added in the private service-providing sectors from December 2007 to February 2020; see US Bureau of Labor Statistics, "Data Retrieval: Employment, Hours, and Earnings (CES)," Table B-1. Employees on Nonfarm Payrolls by Industry Sector and Selected Industry Detail, "All employees, thousands, private service-providing, seasonally adjusted" [CES0800000001]; https://www.bls.gov/webapps/legacy/cesbtab1.htm.

5. Stefania Albanesi and Jiyeon Kim, "Effects of the COVID-19 Recession on the US Labor Market: Occupation, Family, and Gender," *Journal of Economic Perspectives* 35, no. 3 (2021): 3–24.

industries during past recessions, suggesting that these dynamics from previous recessions may start to affect the service industry going forward.

Caroline Hoxby thought it was great to look at labor force participation, employment, and caregiving during the pandemic, but as much of Goldin's work has shown, especially for college-educated women, there are career concerns associated with continuous employment and the degree of engagement in work. Hoxby was interested in whether women were more likely to be distracted workers during the pandemic as a result of multitasking, having to focus also on caring for their children and supervising their schooling, for example. She asked whether women were more likely than men to be disturbed during work by being the person who was primarily responsible for children, de facto if not de jure. Hoxby wondered if, despite their continuous employment and presence, women were ultimately less productive than men because of multitasking. She pondered the long-term effects on women's careers.

Austan Goolsbee responded and said that, interestingly, productivity in most of these sectors went up during the pandemic. He referred to restaurant productivity in the national income accounts where it can be seen that restaurant productivity went up during this period. Goolsbee stated that overall, with real GDP having recovered and the labor force not being back to where it was before the recession, labor productivity for the whole economy is up 4–5 percent, just as much of the service sector. He found it interesting that wages have not yet been reflecting this. Goolsbee also commented that Goldin's results seem to emphasize the difference between the aggregate and the cross section. He referred to work by Furman, Kearney, and Powell, showing the effect of childcare on labor force participation but indicating that it doesn't contribute to the aggregate nearly as much as one may think from looking at the numbers, because one group seemed to be replacing the other group.[6] He wondered, against the backdrop of extreme labor scarcity, if employers substitute by hiring women without children ahead of women with childcare responsibilities. Similarly, he pondered to what extent employers in industries with equal shares of women and men prepandemic were now preferring to hire men and suggested it would be fascinating to look at the issue of the child penalty.

In response to Hoxby, Goldin suggested that working papers in general are leading indicators of where we are headed going forward and predicted

6. Jason Furman, Melissa Schettini Kearney, and Wilson Powell, "The Role of Childcare Challenges in the US Jobs Market Recovery during the COVID-19 Pandemic," working paper 28934 (Cambridge, Mass.: National Bureau of Economic Research, 2021), https://www.nber.org/papers/w28934.

that there is going to be a lot of interesting research on the long-term consequences of these experiences, noting that, unfortunately, terrible times lead to more research.

Olivier Blanchard stated that differences across countries might help in identifying some of these issues. He considered France, which has free childcare, and noted that this is likely to make a difference, observing that France kept schools open throughout the pandemic and had an explicit partial employment system in which people could decrease their work hours. He argued that each of these factors is likely to have implications for what happens to the participation rate of women.

Robert Gordon compared Europe to the United States—the United States lost a lot of jobs and the federal government expenditures were centered on unemployment compensation, whereas in Europe there were central government subsidies for people to keep their jobs and work less. He concluded that this suggests that there was greater job loss in the United States. Turning to the issue of productivity, Gordon drew attention to the huge differences between service sector jobs, which make up about two-thirds of the economy, as indicated by a data set, not yet available online, on productivity differences across industries. He noted that services which involve close contact saw productivity decrease while service jobs which could be done remotely experienced an increase in productivity. He related this to the work of Barrero, Bloom, and Davis, which has shown that increased productivity related to remote work can be partly accounted for by people's commuting time being substituted for work hours rather than leisure time.[7] Gordon noted that—assuming the Bureau of Labor Statistics is not measuring this increase in hours per job—this suggests that some of the increase in productivity of remote work may be a measurement issue. Gordon admitted that none of this relates to men versus women.

Goldin responded to the comments on employment measurements and noted that her paper includes those who had jobs but were not currently at work. Goldin continued, addressing the issue of productivity changes and highlighting the ongoing increase in automation across the service sector, including ordering at restaurants, checking out at the store, and in visits to doctors' offices. Those places have not seen the greatest increases in productivity, she noted, but suggested this may change in the future.

7. Jose Maria Barrero, Nicholas Bloom, and Steven J. Davis, "Why Working from Home Will Stick," working paper 28731 (Cambridge, Mass.: National Bureau of Economic Research, 2021), https://www.nber.org/papers/w28731.

IGOR MAKAROV
London School of Economics

ANTOINETTE SCHOAR
MIT Sloan School of Management

Cryptocurrencies and Decentralized Finance (DeFi)

ABSTRACT The paper provides an overview of cryptocurrencies and decentralized finance (DeFi). The discussion lays out potential benefits and challenges of the new system and presents a comparison to the traditional system of financial intermediation. Our analysis highlights that while the DeFi architecture might have the potential to reduce transaction costs, similar to the traditional financial system, there are several layers where rents can accumulate due to endogenous constraints to competition. We show that the permissionless and pseudonymous design of DeFi generates challenges for enforcing tax compliance and anti–money laundering laws and preventing financial malfeasance. We highlight ways to regulate the DeFi system which would preserve a majority of benefits of the underlying blockchain architecture but support accountability and regulatory compliance.

The financial system performs a wide array of functions that are important for economic growth and stability, such as allocating resources to their most productive use, moving capital from agents with surpluses to those with deficits, and providing efficient means for moving wealth across time and states.[1] To achieve these goals, the US financial system, and similarly most other countries, has traditionally relied on a set of intermediaries such as banks, brokers, and exchanges that are connected by payment systems. These intermediaries serve as centralized nodes that guard the access to

1. See, for example, Merton (1995) or Allen, Carletti, and Gu (2019).

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

Brookings Papers on Economic Activity, Spring 2022: 141–196

the financial system and provide customers with essential services such as record keeping, verification of transactions, settlement, liquidity, and security. This architecture implies that intermediaries perform many of the core functions in the system and also help with the implementation of regulatory goals such as tax reporting, anti–money laundering laws, and consumer financial protection. As a result, however, these intermediaries can hold significant power, based on their preferential access to customers and data. This centralized position, if not properly harnessed and regulated, can be a source of outsized economic rents and can lead to considerable inefficiencies. It can also lead to inherent fragility and systemic risk if core intermediaries become corrupted or investors lose trust in the system.

The concern about the power and potential corruptibility or fragility of intermediaries, possibly heightened by the experience of the 2008 financial crisis, has contributed to the new "revolution" brought about by blockchain technology, which is one of the fastest growing financial innovations over the last decade. Its attraction lies in the ability to build decentralized and open access platforms that reduce the reliance on centralized trusted intermediaries and middlemen.

Eliminating unnecessary intermediaries can potentially be a significant benefit of blockchain architecture. Technological innovations have, of course, long been consequential in improving the efficiency of the financial system or strengthening competition. We can think of innovations like mobile banking and algorithmic lending. What differentiates blockchain from past technological innovations is that it offers the possibility of a completely different financial architecture, commonly called decentralized finance (DeFi), where record keeping is decentralized, access to the system is anonymous and unrestricted, and any form of intermediation would be built on top of it.[2]

To assess the potential benefits and challenges of the proposed new architecture, it is important to recognize that intermediaries are not merely gatekeepers which have no economic value except for rent extraction. Many problems with existing intermediaries originate from the economic forces that are an inherent part of financial markets and therefore exist also in DeFi solutions but might be relocated to different layers in the new infrastructure, as we will discuss. In addition, some of the rents that financial

2. DeFi is also distinct from the generic umbrella term *fintech*. While fintech innovations also introduce new technologies to financial services—for example, Rocket Mortgage, which uses online origination in mortgage lending—they still rely on a model of centralized intermediaries.

institutions enjoy in the current financial system are a deliberate regulatory choice: in order to provide institutions with the incentives to abide by regulations, rule makers allow these institutions to earn some rents to ensure that they have a franchise value.

Advocates of DeFi solutions argue that financial services are ripe to undergo dramatic and disruptive changes. How this evolves, in terms of technology, regulation, and ultimately liquidity and credit to the economy, has important consequences for the United States and global economies. There are also strategic and competitive implications across countries. The goal of this paper is to raise some of the issues that arise in a system of decentralized finance and propose some solutions, while at the same time providing an introduction to how such a system works and the mechanics behind it.

We start by laying out how the blockchain technology that underpins virtually all DeFi solutions works. We discuss the different ways security is achieved under different protocols, in particular proof of work (PoW) and proof of stake (PoS), and what economic incentives are built into these solutions to ensure the integrity of the blockchain ledger. Our analysis highlights that the current security protocols have built-in economic incentives for concentration of mining or validator capacity due to inherent fixed costs and benefits of coinsurance for validators. We also show that large PoW networks can have negative externalities on the security of smaller PoW networks, which has important implications for the competitiveness of PoW protocols. For PoS platforms, an added complexity arises from the fact that the going concern value of the platform also affects the security of the platform itself and applications that run on it.

Next, we discuss the benefits and limitations of smart contracts. These are self-executing pieces of scripting code that can in theory carry out any computation and are the building blocks of many DeFi applications. Since smart contracts are designed not to have recourse to the legal system, they have to be written as complete contracts up front. We highlight the implications of such a change on the enforcement of contracts, the transaction costs of writing contracts, the opportunity of opting out of current remedial laws, and challenges for consumer financial protection if smart contracts are written outside typical legal protections. Many of these challenges might give rise to a new layer of "trusted" intermediaries, in particular, coders who will help people to navigate DeFi infrastructure that might be too complicated for individual participants. In this context, we explain the role and design of oracles, which provide access to data from outside the blockchain and allow smart contracts to interact with the real world. Based

on these building blocks, we then provide an overview of the current crypto landscape and the main DeFi applications, such as decentralized crypto exchanges, borrowing and lending markets, and yield farming.

Finally, we compare this new DeFi architecture to traditional financial market solutions and lay out how these two regimes solve some of the most important problems in financial systems, such as data privacy and transparency, extraction of rents, transactions costs, governance issues, and systemic risk.[3]

DeFi applications might have the potential to democratize finance by creating a level playing field among providers of financial products and services. But we show that the current design of DeFi applications, which are predominantly built on permissionless and pseudonymous blockchains, generates formidable challenges for tax enforcement, aggravates issues of money laundering and other kinds of financial malfeasance, and, as a result, creates negative externalities on the rest of the economy. Similar to the traditional financial system, there are several natural points where rents can accumulate at different layers in DeFi architecture due to endogenous constraints to competition caused by network externalities and economies of scale. Also, rent extraction can be driven by frictions at the customer level due to lack of financial sophistication or behavioral biases. In cases where market competition does not work to restrict excessive rents, regulations are typically established to protect the interest of users. But here again, the permissionless and pseudonymous design severely limits the ability of regulators to restrict unscrupulous operators.

The pseudonymous and permissionless structure also has implications for the governance of DeFi apps. Many DeFi apps, in their quest to avoid placing trust in any actor or institution, have experimented with new organizational forms, so-called decentralized autonomous organization (DAO). The basic idea of DAO is to spread control over decisions among all interested stakeholders by issuing special "governance" tokens that give their holders the power to propose changes to the protocol and vote on them. We discuss the governance challenges that arise in such arrangements and show that they face the same fundamental governance issues as traditional organizations. As a result, we show that in the majority of crypto projects ownership is concentrated.

Lastly, we discuss the potential of DeFi solutions to contribute to systemic risk and have spillover effects on the rest of the economy. We

3. Harvey, Ramachandran, and Santoro (2021), Schär (2021), and Aramonte, Huang, and Schrimpf (2021) also provide detailed discussions of the DeFi ecosystem and its applications.

highlight that DeFi so far has operated under a narrow banking model. This removes many of the problems faced by the fractional reserve system but also constrains the efficient use of capital. Presently, the main systemic risk comes from the ability of investors to take highly leveraged and interconnected positions and a potential run on stablecoins. So far, the systemic risk has been limited, but as ties between the regular financial system and DeFi increase, the risk can grow.

We conclude by discussing challenges and potential solutions for regulators and market participants in this new infrastructure. A natural place for regulatory oversight in this new ecosystem is at the level of developers and validators, who in turn control the network protocol. Once this level of regulatory compliance is established, many other functions can be built that would address the majority of issues we outlined above. This solution looks similar to a permissioned blockchain, but it preserves most of the desired properties of the blockchain such as observability of transactions, automatic settlement, and execution of the same set of smart contracts.

If regulators give up on the ability to oversee validators, the effectiveness of regulation will be much more limited and will depend on the goodwill and voluntary cooperation of validators and developers of the blockchain. If validators accept transactions from every party, the most regulators could hope for is to separate the network into regulated and unregulated parts. The latter part could then harbor bad actors and facilitate illegal activities. The opportunities of sidestepping the regulated part will generally increase with the level of crypto adoption, since people will be able to transact predominantly in the unregulated part and avoid triggering regulatory oversight.

I. Blockchain Technology

A typical financial system can be represented, at an abstract level, as a collection of states and transactions that describe the transition from one state to another. For example, in a payment system a state is a collection of all the accounts in the system together with their balances. Transactions specify how funds move between accounts.

Historically, financial intermediaries have been the key nodes in the financial system that control the accuracy of customer accounts, perform bookkeeping functions, and ensure that unauthorized persons do not have access to an account. For a long time, this centralized model of bookkeeping was the only viable option. But recent advances in technology have enabled an alternative architecture of storing and managing information where no

single entity has full control over all the states and transactions or any subset of them. Instead, multiple parties (validators) hold their own copies of states and jointly decide which transactions are admissible. This architecture became known as distributed ledger technology (DLT). A blockchain is a form of DLT in which all transactions are recorded and organized in blocks that are linked together using cryptography. Bitcoin was the first and remains the most famous application of blockchain technology.

One of the main advantages of DLT is the elimination of a central point of failure. Since multiple copies of records exist, the corruption of a single node or a single copy has no effect on the security of the blockchain. In fact, blockchain protocol allows for multiple points of failure or corruption as long as the majority of validators are not corrupted. In particular, it allows validators to be parties that do not trust one another or are even adversaries.

Blockchains are usually divided into permissioned and permissionless ledgers depending on the set of entities that are allowed to be validators. In a permissioned blockchain, a set of validators is approved by a coordinating body, which can be a private firm or a consortium of institutions. In contrast, a permissionless blockchain does not impose ex ante constraints on the number or identity of validators. In addition, blockchains are sometimes categorized as private or public ledgers. In a public blockchain, everyone has full access to the information stored on the blockchain. In contrast, only authorized parties can observe transactions in private blockchains. Typically, permissioned blockchains are private, and permissionless blockchains are public.

Permissioned blockchains still require trust in the coordinating body that approves validators, which is viewed by many crypto enthusiasts as a fundamental flaw. In contrast, permissionless blockchains do not rely on trust in any individual validator, forming what famously has been called a "trustless" trust architecture. The trustless trust, however, comes at a high cost. Since anyone can become a validator in a permissionless blockchain, the system is potentially vulnerable to a Sybil attack where an adversary subverts the system by creating a large number of pseudonymous validators and uses them to gain disproportionately large influence over the consensus protocol.

Two main approaches have been proposed for permissionless protocols to be resilient to a Sybil attack: proof of work (PoW) and proof of stake (PoS). The main idea behind both approaches for validating transactions is to provide validators with a reward for their services and to make it costly for an adversary to attain a majority stake and subvert the system. The

reward is meant to provide validators with financial incentives to work honestly. The reward usually combines two parts: transaction fees and a prespecified amount also known as a block reward. The block reward is typically denominated in the platform's native currency and is financed through issuance of new coins, thus serving as a dilution tax on all users.

The decentralization of the ledger also has implications for the scalability of the network. Intuitively, as the ledger becomes more decentralized more copies need to be distributed and more resources need to be spent to achieve the protocol consensus and make the blockchain secure. This trade-off between decentralization, security, and scalability was famously formulated by Vitalik Buterin, a cofounder of Ethereum, in the early days of Ethereum and became known as the scalability trilemma (or sometimes as the blockchain trilemma). The trilemma has attracted a lot of attention, and a large number of new blockchain solutions have been introduced to achieve the three goals simultaneously.[4]

In the following, we leave aside the technical issues such as scalability. We also refrain from a game theory analysis of security of different protocols.[5] Instead, we focus on the embedded economic mechanisms and incentives that are at the heart of the different protocol security approaches. Since most DeFi applications are currently built on permissionless blockchains, we will focus predominantly on these blockchains. We show that both PoW and PoS favor validator concentration, since there are strong implicit incentives for validators to pool their capacity and coinsure their risk of winning a block reward. We also discuss the resilience of PoW and PoS to an attack and show that large existing networks have negative externality on small networks. These properties have important implications for competition in the crypto space, which we discuss in section IV.

I.A. PoW Protocols

In a PoW protocol such as Bitcoin, validators (also known as miners) compete for the right to verify transactions and obtain their reward by solving a computationally intensive problem. For a successful attack on a blockchain an attacker needs to control a large fraction of the total network power, typically 51 percent, which resulted in the nickname "51% attack." Once an attacker controls the majority of mining power they can alter

4. These include sharding, sidechains, and lightning networks. There are also non-blockchain solutions, for example, hashgraph technologies.

5. For an example of such analysis, see Biais and others (2019) and Halaburda, He, and Li (2021).

transactions in the system, for example, they can spend the same cryptocurrency multiple times (known as a double-spending attack).

The likelihood of an attack in a PoW protocol therefore depends on the prospects that a malevolent party amasses enough computing power. Notice that miners should at least break even in the long run to be willing to invest in mining. Thus the expected rewards collected for mining a block should cover the cost of its mining. This implies that there are no economic disincentives of amassing 51 percent and the constraint is on the feasibility of amassing 51 percent of hashing power (Budish 2018).[6]

Of course, any successful attack on a blockchain reduces trust in this blockchain and therefore its economic value. If miners have to incur large fixed costs to set up their operations, then by attacking the blockchain they will forfeit some of the future profits and might not be able to recover their initial investments. This reduces the benefits of the attack and can make it unprofitable.

The lower the fixed costs, the less costly is a 51% attack. As a result, any factors that reduce fixed costs have negative effects on the security of the network. In particular, large PoW networks like Bitcoin or Ethereum have negative externalities on the security of smaller PoW networks.

The large appreciation of Bitcoin and Ethereum led to significant investments in mining capacity.[7] Smaller networks like Litecoin or Bitcoin Gold usually attract only a small fraction of the mining capacity of these larger coins, since their rewards also are much lower. This creates a possibility that a miner with a large hashing capacity can divert a fraction of it to attack a smaller coin, if they chose to.

Furthermore, the emergence of market places like NiceHash, where mining hash power can be rented for a specific time period, has made it possible for people to speculate on mining profitability without owning the physical hardware themselves and to amass hashing power for a possible attack. The amount of available hashing power in these marketplaces is only a small fraction of the capacity used in large networks such as Bitcoin and Ethereum, which usually operate close to full capacity. But the available capacity on NiceHash often is significantly larger than the total mining

6. Hashing power or hashrate is the amount of computer power that a network consumes to operate; see BitDegree, "What Is Hash Power (Hashrate)?," https://www.bitdegree.org/crypto/learn/crypto-terms/what-is-hash-power-hashrate.

7. The global mining capacity of BTC increased more than one hundredfold and ETH more than three hundredfold over the last four years; see CoinWarz, "Bitcoin Hashrate Chart," https://www.coinwarz.com/mining/bitcoin/hashrate-chart.

capacity employed in smaller networks.[8] These renting opportunities have significantly reduced the cost of a 51% attack on smaller networks and in fact have led to many such attacks on smaller cryptocurrencies such as Bitcoin SV, Bitcoin Gold, and Ethereum Classic (see table A.1 in the online appendix).

The negative externalities of large PoW networks on smaller networks have important implications for the competitiveness of PoW protocols. It suggests that once one or a few major PoW blockchains are in existence, new entrants might find it difficult to compete. While the new protocol has not reached a critical mass yet, it has a heightened likelihood of being subject to an attack. This makes it less secure and might reinforce the dominant position of the first movers. One defense against the negative externalities of hashing capacity in larger blockchains would be to make mining equipment very platform specific, so that slack in a larger system does not affect the new entrant. However, platform-specific mining hardware can increase entry cost for miners to the new platform, which can have a negative effect on its growth and security.

While there have not been any successful 51% attacks on Bitcoin or Ethereum, this does not mean they are completely safe from them. First, as we mentioned above, these networks have benefited so far from large price appreciation that have made miners operate at nearly full capacity. If at some point there is a substantial price decline, it is likely that an increasing number of miners will find it unprofitable to continue their mining operations. This can lead to an increase in spare mining renting capacity and might increase the probability of an attack.

Second, in the original design, Satoshi Nakamoto, the inventor of Bitcoin, envisioned a world where mining would be fully decentralized and not depend on a few large players. In this world, miners would find it difficult to collude, and failure of any one miner would have no consequence for the security of the network.

This original idea, however, clashes with the economics of mining in PoW protocols. By design, the probability of winning the race and obtaining the block reward is proportional to the computing power spent on mining. This gives strong incentives for miners to pool their computing power and coinsure each other. As a result, mining in most PoW blockchains is dominated by large mining pools (Cong, He, and Li 2021; Ferreira, Li, and Nikolowa 2019).

8. See, for example, the website Crypto51. https://www.crypto51.app/. which measures the cost to 51% attack Bitcoin and other major PoW cryptocurrencies.

The concentration of mining pools has attracted a lot of public attention and concern, since high concentration facilitates collusion among miners and, with it, the danger of an attack. Even if miners themselves do not misbehave, high concentration increases the risk that a malevolent party, either a private or a state actor, could hijack them and gain control over the network.

Some observers downplayed the risk of the attack coming from pool concentration, arguing that even though pools can have substantial influence over the cryptocurrency protocol, they do not necessarily control their miners. Therefore, if any pool is noticed engaging in rogue behavior, its miners can leave it and join other pools.

The power that a pool operator has vis-à-vis individual miners depends on the ease with which miners can shift capacity across pools, which in turn depends on the underlying size distribution of the miners. In Makarov and Schoar (2021) we document that miner concentration in the Bitcoin protocol is high, even at the level of individual miners. We show that, at times, fewer than fifty miners control 50 percent of mining capacity. One explanation for this concentration in mining power seems to lie with the high fixed costs of setting up a large mining farm that result in increasing returns to scale.

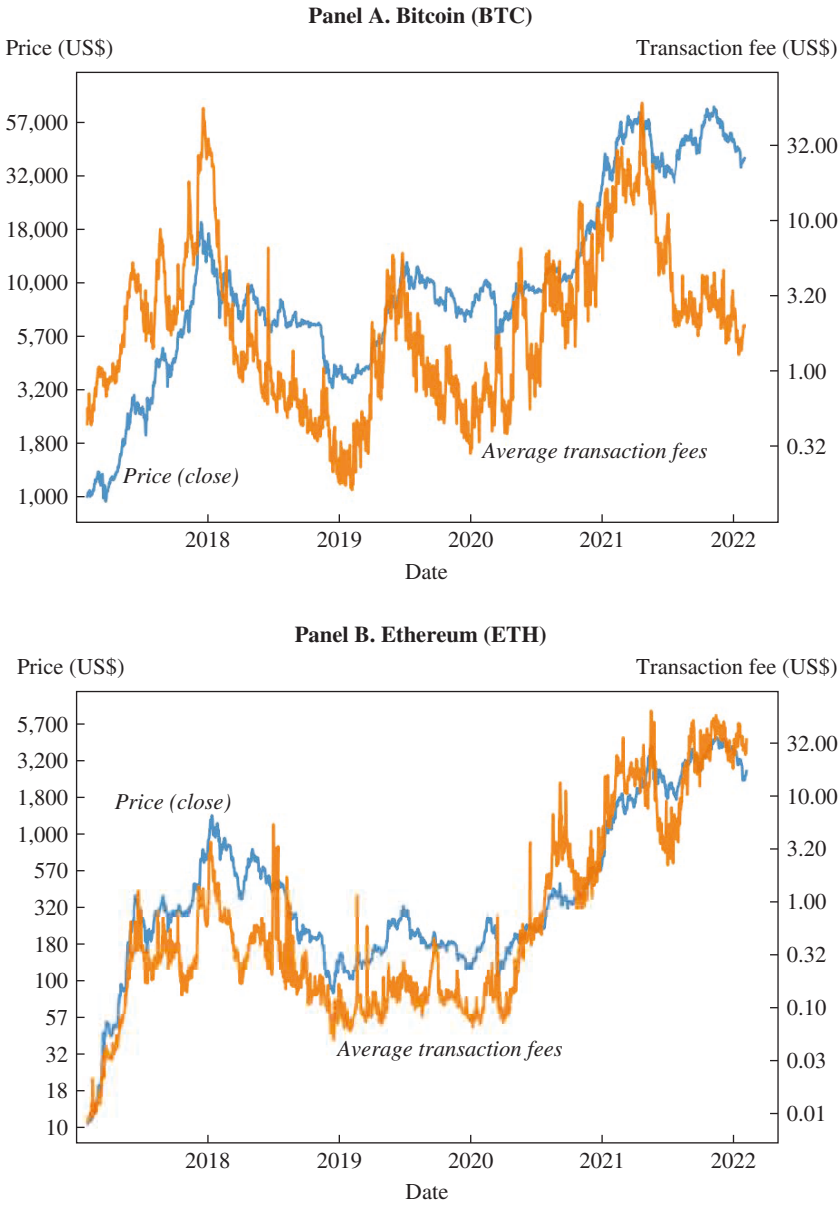
The paper also shows that the concentration of mining capacity is countercyclical and varies with the Bitcoin price. It decreases following sharp increases in the Bitcoin price and increases in periods when the price drops. Thus, the risk of a 51% attack increases when the Bitcoin price drops and makes the system more fragile.

I.B. PoS Protocols

While the costs of an attack and the resilience of a PoW network increase with the size of the network, so does the cost of verification. According to the Cambridge Bitcoin Electricity Consumption Index, the annual electricity consumption of the Bitcoin network in 2021 reached 130 TWh, which exceeds the annual consumption of such countries as Norway or Ukraine. Because miners have to be compensated for their costs, large electricity consumption translates into high transaction fees. Figure 1 shows the average transaction fees in the two largest PoW protocols, Bitcoin and Ethereum. As the Bitcoin and Ethereum prices have significantly increased over time, so have the fees.

The serious concerns about the sustainability and energy consumption of PoW protocols have favored the emergence of PoS blockchains. PoS protocols consume significantly fewer resources than PoW protocols. Platt

Figure 1. Average Transaction Fee and Price for Bitcoin and Ethereum



Sources: Messari.io and authors' calculations.

Note: This figure shows the daily transaction fees and closing prices for Bitcoin and Ethereum from January 2017 to February 2022. Daily closing prices are plotted on the left axis and daily average transaction fees are plotted on the right axis. The figures are plotted in log scale.

and others (2021) estimate energy consumption of major PoS protocols and show that their energy consumption per transaction is comparable to that in the Visa network. Recognizing the drawbacks of PoW protocols, after 2017 there was a significant acceleration in the development of PoS blockchains. Also, Ethereum instituted a shift to a PoS protocol, Ethereum 2.0, to be completed in 2022.

In a PoS protocol, instead of solving a difficult mathematical problem, a validator stakes its coins, which can be forfeited if the validator fails to verify transactions in a timely manner or its actions are determined to be malicious. In most PoS protocols, participants who stake more coins are more likely to be chosen to verify transactions (or have more rights to vote for a validator in delegated PoS networks). Thus, PoS protocols are built on the idea that a party that has a large stake in the given network would not want to undermine this network since the gains from an attack would not compensate for the loss of value that comes from penalties and the drop in the network's valuation.

The above argument relies on the idea that a validator which owns a large stake in the platform also has an interest in its continuation value and thus should be disincentivized from endangering it. This logic makes sense, if the attack in question is, for example, a double-spending attack, since the gains in that case are a small fraction of the total value of the network.

However, the gains from an attack might not be restricted to simple gains from double-spending. First, if the network is part of a competitive environment, competing networks might realize substantial gains from undermining a new entrant. Similar to what we described in PoW blockchains, undermining fledgling rivals can be particularly profitable if it reduces future competition.

Second, many PoS blockchains are smart contract platforms that position themselves as a base layer providing security for other applications or even other blockchains that are built on it. In this case, there is tension between the value of the base layer blockchain and its applications. If the value of the base layer is below the value of an application, an attacker who wants to undermine the application might find it profitable to attack the base layer. To prevent such an attack, the value of the blockchain at the base layer should be substantially greater than the value of its applications. Since the value of the base layer comes primarily from transaction fees (and seigniorage), the possibility of an attack on the base layer puts a lower bound on the required size of the fees that have to accrue to the blockchain at the base layer. High fees, however, hurt the value of applications built on the platform, and thus the platform's value.

Table 1. Concentration of Validator Stakes

Cryptocurrency	*Amount staked (% of circulating supply)*	*Validator concentration (%)*	
		Top 10	*Top 50*
Solana	70	23	56
Cardano	73	30	47
Avalanche	97	17	57
Terra	77	36	76
Polkadot	57	30	56
Cosmos Hub	63	45	87
NEAR Protocol	61	50	96
Polygon	34	72	99
Fantom	54	88	100
Tezos	76	63	96

Sources: Stakingrewards.com and authors' calculations.

Note: This table reports the concentration of validator stakes for the top ten proof-of-stake smart contract platforms by market capitalization as of February 2022. Validator stakes include stakes provided by validators themselves and stakes delegated to validators. The data exclude Ethereum since it is in a transition period.

We showed in section I.A that mining in PoW blockchains is dominated by pools because they allow miners to coinsure each other. A similar force is at play in PoS blockchains. Since the probability of being chosen and collecting the reward depends on the amount of coins a validator is staking, investors have incentives to pool their stakes together and coinsure each other.

Table 1 documents concentration of validators for the largest PoS protocols as of February 2022. The data show significant concentration for the vast majority of the PoS blockchains. The top ten validators hold typically more than 25 percent of the capacity, while the top fifty validators are above 50 percent.

In addition, since the technology used across different PoS protocols shares many similarities, the same validators typically work on multiple blockchains. Table 2 shows the top fifteen validators together with their combined stakes in the top ten largest PoS protocols. The top ten, fifty, and one hundred validators account for 14 percent, 32 percent, and 41 percent of stakes across the ten largest PoS blockchains, respectively.[9]

The concentration of PoS validators at the time of writing is lower than in the PoW protocols, but it is not fully dispersed either. It is of interest that a few validators are starting to emerge as dominant players across different blockchains.

9. Authors' calculations and data from Stakingrewards.com.

Table 2. Top Validators

Validator	*Staked (US$ billions)*	*Share (%)*
Everstake	2.8	2.2
Binance Staking	2.6	2.1
Chorus One	1.6	1.3
Dokia Capital	1.6	1.3
Certus One	1.5	1.2
Bison Trails	1.5	1.2
Allnodes	1.5	1.2
InfStones	1.5	1.2
Kraken	1.4	1.1
Staked	1.2	1.0
P2P Validator	1.2	1.0
Orion Money	1.1	0.9
B-Harvest	1.0	0.8
Staking Facilities	1.0	0.8
Figment	1.0	0.8

Sources: Stakingrewards.com and authors' calculations.

Note: This table reports the top fifteen proof-of-stake validators and their aggregate stakes in the top ten proof-of-stake smart contract platforms by market capitalization as of February 2022. Validator stakes include stakes provided by validators themselves and stakes delegated to validators.

II. Smart Contracts

Smart contracts have become another fundamental layer of the new DeFi architecture. To go beyond simple interactions such as the transfer of coins or assets on the blockchain, many newer protocols starting from Ethereum provide the opportunity to embed pieces of scripting code that can, in theory, carry out any computation. These pieces of code became known as "smart contracts." The term and the concept are credited to the cryptographer Nick Szabo, who defined smart contracts as "a set of promises, specified in digital form, including protocols within which the parties perform on these promises" (Szabo 1996, par. 5). The modern implementation of this idea arrived with the creation of Ethereum, which is designed to execute smart contracts and make it convenient for developers to build applications on top of the blockchain.[10]

By itself, using software code to represent and execute contractual agreements is not new. For example, when trading via an online brokerage platform, each time a customer sets up a limit order that automatically buys certain stocks when prices match a predefined level, the contract is

10. See Buterin (2014).

executed by a software program. Financial markets and e-commerce are dominated by these types of arrangements since they allow a large volume of transactions to be executed quickly and efficiently. But even if the program automatically executes a set of tasks, in traditional electronic contracts, the parties to the contract still have recourse to the legal system if there is a dispute. For example, if a limit order is executed based on wrong information used by the online brokerage platform, the client can seek restitution from the brokerage through the courts.

The critical differences, from an economic perspective, between traditional electronic arrangements and smart contracts that are executed on a permissionless blockchain arise from how the contracts are executed and enforced.[11] We show that since smart contracts are self-executing once they have been embedded in the blockchain, they require contracting parties to complete contracts as much as possible ex ante, since they cannot rely on ex post remedial protections through the legal system. We discuss the implications of this switch for the transaction costs of writing contracts, the ability of contracting parties to opt out of the current legal protections, and the constraints to consumer financial protections. The need to import up-to-date information from the outside (off-chain) world into the blockchain also led to the development of a new set of entities, so-called oracles. We lay out the role of oracles for the functioning of smart contracts and potential vulnerabilities that are introduced through oracles. Finally, we argue that this new architecture might require contracting parties to rely on a new set of trusted intermediaries, such as the developers of the smart contract platform or coders who help to write the computer programs that will be executed on the blockchain.

II.A. Execution and Enforcement

The execution of a smart contract on a permissionless blockchain fundamentally changes the process of enforcement (Werbach and Cornell 2017; Werbach 2018). First, once a program has been executed, the distributed nature of the contract verification makes it impossible to unilaterally stop or reverse its execution, unless certain conditions for stopping the smart contract were included in the program ex ante. Second, even if one party wanted to sue a counterparty, there might not be any party that can be

11. Smart contracts can also be implemented on permissioned blockchains. In this paper, we focus on smart contracts run on permissionless and public blockchain protocols, since their major applications have been hosted on such blockchains.

held accountable because of the anonymity of the transactions. Practically speaking, there might be no one who can be served with a legal notice.

These changes are important for the application of contract law, since it is fundamentally a remedial institution that operates on an ex post basis. First, contract law aims to rectify situations ex post, where one party has wronged another party by breaching the terms of the contract or not delivering on a promised action. Second, the law incorporates a variety of doctrines which allow one or multiple parties to annul the contract ex post. These exemptions are meant to protect contracting parties against unwittingly (or deliberately) taking advantage of each other or of an unforeseen situation. These are issues such as unconscionability, mutual mistake, illegality, capacity, consideration, fraud, or duress. The role of judges and the legal system is to oversee and enforce the intended application of the law in these cases. In other words, the legal system completes contracts that were either deliberately or unintentionally left incomplete ex ante (Wright and De Filippi 2015).

Of course, contracts are written in the shadow of the law. The expectations that contracting parties have about how laws will be enforced affect how contracts are written in the first place and which parts can be left unspecified. Since smart contracts do not allow for recourse to the legal system, they have to be written as complete contracts up front. Or, at a minimum, the contracting parties have to specify exactly which states of the world they are willing to leave unspecified. Since the smart contract cannot be unilaterally stopped and renegotiated, if a state of the world is not ex ante specified, the program will execute as if this state never existed.

This highlights that a contract breach in the traditional sense is not possible on the blockchain. Once the parameters encoded in the smart contracts are realized, the code will execute the transaction. This significantly reduces the chance of one party to a contract reneging on it after the fact, say, because they changed their mind or they were not serious about the transaction in the first place. But the automatic execution of smart contracts also eliminates the opportunity for "efficient breach." Take the situation of a mutual mistake: a buyer and seller agree to the purchase of an asset at a specific price, but just before the seller is supposed to deliver the asset, the seller discovers that the asset is worth much more than either side had realized. Here, in a traditional contracting situation, the seller could engage in efficient breach and not deliver the asset until both sides had a chance to renegotiate the terms of the deal. However, with a smart contract, the transfer will be executed since the parties by definition did not plan for the mutual

mistake up front. A similar logic holds for many of the other protections that traditional contract law provides. This shifts the status quo of which party will be in the role of plaintiff and defendant.[12]

II.B. Smart Contract Trade-Offs

TRANSACTION COSTS OF CONTRACTING As the discussion above highlights, smart contracts must be written in precise, fully defined computer code since they cannot be modified once executed. Many proponents of smart contracts have suggested that this reduces their cost since there is no scope for ex post renegotiation. But these cost savings might be offset by the higher up-front costs of negotiating and specifying the precise terms of an agreement in all possible states of the world. These up-front costs will become especially high when there is large uncertainty about the future states of the world or if these states are hard to imagine and to define ex ante.

To mitigate these issues, traditional contract law systems provide a series of mandatory and default positions that allocate risk when matters are left unspecified. In the case of smart contracts this recourse to the legal system is not possible. So the costs must be borne by the individuals engaging in the contract. In the case of contracts that are very simple and standardizable, some templates of code will most likely be developed which anyone can use to embed in a smart contract. This can reduce the up-front cost in cases where many people have very similar contract issues and the future states and outcomes over which the contract needs to be defined are also very standard and simple to understand. However, as soon as there is more variation in possible contract templates to be considered in a contracting situation, the mental cost of comparing and understanding the different options might become quite high. And of course, the costs are even higher if the situation is unique and a lot of value is at stake. Here parties cannot choose from existing templates but have a strong incentive to not inadvertently miss or miscode a possible state of the world. This means they do have to bear the up-front costs of trying to write as complete a contract as possible.

SMART CONTRACTS AS A COMMITMENT DEVICE Even people who trust the legal system might in some situations want to avoid ex post litigation risk to bring down ex ante cost, for example, reducing the possibility of

12. Parties to a smart contract could try putting in protections against mutual mistakes by writing into the contract arbitration of third-party experts, but this would require trust in experts and therefore, would go against the main idea of smart contracts.

opportunistic behavior or efficient contract breach ex post. Take a situation where both parties to a contract are well informed about the functioning of a certain financial product, say, a mortgage, and thus ideally the lender would not need to spend time developing education material to inform the borrower about what happens in case of default. However, if the borrower has the right to sue ex post if they were not informed that the lender can seize the property, the lender will be forced to develop training material to prove that the borrower has been informed. An informed borrower and lender might be better off if they could shut off the opportunity for the borrower to sue in case of default. It would eliminate the lender's need to invest in expensive training material which is wasteful in this case. But since the borrower cannot abdicate their right to sue, both parties must bear the cost of the up-front training.

These issues apply in situations where both parties to a contract are sure that they do not value any ex post protection through contract laws. This requires that both sides must be well informed about the logic of the contract and all the possible ex post outcomes and do not fear the possibility of being taken advantage of. In financial markets this is an important concern since many contracts involve investments in complex and risky products, for example, trading in derivatives. If customers could sue each time a bad state of the world occurs and claim that they were misled about the product, intermediaries would not be able to sell any risky securities. In the United States the law has addressed these issues by granting certain exemptions to high-net-worth individuals or people who can demonstrate their knowledge in those products. But it does not provide sweeping exemptions from the ex post protections of contract law since in many situations consumers might not even be aware of their own lack of knowledge relative to an informed market participant.

SMART CONTRACTS AND CONSUMER FINANCIAL PROTECTION A large body of literature in finance has shown that many participants in financial contracts, especially retail investors, lack financial literacy and are not well prepared to understand financial markets.[13] Although parties are generally free to enter into agreements, subject to certain limitations and exceptions, the law protects parties in certain situations by determining whether they had the capacity to enter into a legally binding agreement. For example, contracts may be voidable if made by a minor or persons who are mentally ill or intoxicated at the time of contracting. By not allowing mandatory

13. See, for example, Lusardi and Mitchell (2007).

ex post protections through the legal system, smart contracts do not provide sufficient safeguards for financially less informed or more fragile customers. Since smart contracts typically have limited means to test for a person's financial sophistication or mental capacity, the enforcement of these contracts could lead to undesirable outcomes if there is no provision to reverse the outcome as in traditional contract law.

If financially less sophisticated consumers are aware of their lack of knowledge and understand that there is a risk that in such an environment they are disadvantaged, the most plausible result would be to opt out of this contracting environment. However, if smart contracts became the predominant form of contracting, it would severely affect market participation of less sophisticated consumers. Or, alternatively, these customers would have to find trusted intermediaries to act on their behalf. So we are back to the original problem of how to ensure good performance of intermediaries. But given the pseudonymity of the blockchain environment, it would be more difficult to build trust. Furthermore, a large body of literature in behavioral finance has shown that many financially unsophisticated consumers are not aware of their lack of information or are overoptimistic about their ability to participate in financial markets. As a result they might unknowingly sign contracts that are against their own interests.[14]

To curtail the most egregious abuses in the traditional system, the United States has a set of consumer financial protection regulations in place, including the Consumer Financial Protection Act, the Fair Debt Collection Practices Act, and the Truth in Lending Act. These aim to reduce the asymmetry in knowledge and information between financial institutions and customers to provide better outcomes for consumers. As the discussion of smart contracts suggests, these types of regulations will be difficult to implement on a permissionless blockchain.

ARE SMART CONTRACTS REALLY "TRUSTLESS"? An often highlighted promise of smart contracts is that they may reduce the need for trust between contracting parties or trust in the legal system. Legal enforcement of contracts can be cumbersome and prone to error. In some societies the legal system itself can even be corrupt and biased. If people do not trust the legal system, they might prefer a decentralized execution that is not subject to ex post discretion. But it is not clear whether trust can be removed altogether from the process of smart contracting or whether it simply requires a shift of trust to other intermediaries and systems.

14. See, for example, Laibson, Repetto, and Tobacman (2007) or Campbell (2016).

In a narrow set of circumstances, smart contracts can automatically enforce transactions if all parts of the transaction are on-chain. For example, a contract that exchanges one token for another on the same blockchain does not rely on enforcement or adjudication outside the blockchain. Here the level of trust is as high as the trust in the blockchain itself, but some level of trust is still required. For example, parties need to trust the developers who oversee a network's protocol not to have embedded errors in the coding of the platform or that the consensus protocol is well enough designed that it is not prone to any attacks.

However, the vast majority of important financial interactions rely on assets, actions, or information that exist outside the blockchain. For example, one of the most important financial contracts a typical household in the United States makes is for a mortgage against their house. While one could imagine a smart contract that uses the home as collateral, the transfer of the house cannot be fully automated on the blockchain ledger. First, the smart contract would have to stipulate how the deed record in the public database must change, in case of default or non-repayment of the loan. Second, even if we assume that the deed record itself lives on the same blockchain, if the person who currently occupies the house does not move out when the ownership changes, it does need off-chain verification and enforcement to change the de facto state that matters, for example, can you occupy the house you supposedly own.

Getting off-chain data presents a number of challenges. The solution revolves around the use of an oracle—an off-chain entity that creates a transaction on-chain with the data posted. Oracles define how a smart contract incorporates off-chain information into the execution of a program, which we discuss in detail in section II.C. The consequence of using oracles is that parties need to trust them.

In addition, given the lack of an ex post appeals process via the law, a lot is at stake when specifying a smart contract to be as complete as possible up front. Especially for transactions that are more complicated, the machine-readable code for the smart contract must be complete and follow strict rules of syntax and semantics. In practice, most people are not able to write this type of contract themselves and therefore must rely on coders or third-party developers. This can lead to perverse incentive for developers who are more knowledgeable than the principal who hires them to take advantage of the principal and exploit deliberate vulnerabilities in the code. The fact that the code underlying the contract is stored on the blockchain and publicly accessible alleviates but does not completely eliminate the problem. The pseudonymity of the blockchain makes it difficult to

confirm if the developer of a code is also the agent benefiting from any vulnerability. And at least currently, developers are not bound by the same fiduciary standards as financial intermediaries.

OBSERVABILITY When interacting with a regular server-based web application, the user often cannot observe the details of the application's internal logic. As a result, the user has to trust the application service provider. Smart contracts mitigate this problem and ensure that an application runs as expected, since the code underlying the contract is stored on the blockchain and publicly accessible. However, this type of observability can also have a downside if it leads to strategic behavior. For example, take any rating system in finance such as a personal credit score or a firm's bond rating. If the smart contract spells out exactly how the score is calculated, users might optimize against the code so that they land just above the cutoff for the best category. This could undermine the usefulness of these types of scores.[15]

Another possible problem with the observability of data on the blockchain has been highlighted in Cong and He (2019). Since generating decentralized consensus entails distributing information, it changes the information environment for the market participants. In particular, as Cong and He (2019) argue, it can encourage greater collusion between interested parties.

II.C. Oracles

While the blockchain tries to remove the reliance on third-party enforcement, smart contracts often need to access data from outside the blockchain if they want to interact with the real world. Consider, for example, a limit order, where a person writes a smart contract to automatically sell a token of Bitcoin when the price hits a certain target level. For this contract to work, the contract needs to access up-to-date Bitcoin prices. If the data are not obtained in an accurate and timely fashion, a smart trader could reap large gains by taking advantage of stale or wrong prices.

One solution would have been to allow the smart contract to obtain the price by querying an application programming interface (API) of some exchange. The problem with this solution is that almost all blockchains are designed to be deterministic, which means that any state should be reproducible given the history of the network transactions. Determinism is important so that different nodes that execute the contract can come to a consensus. Since querying the internet can, in general, produce different

15. See Berg, Puri, and Rocholl (2020) for an example of loan officers gaming a scoring threshold.

values (for example, the price depends on the time of the query), allowing the smart contract to query the price would lead to different values across the nodes, thus making the consensus impossible.

A solution to the above problem is to use an off-chain entity that does the query and posts the data on-chain. Once the data are on-chain, smart contracts can access and use them. The off-chain entities that query, verify, and authenticate external data sources and then transmit the information to a blockchain, in the crypto parlance, are called oracles.

There are many types of oracles.[16] The central issue in the design of any oracle is trust. Similar to a chain, which is as strong as its weakest link, a smart contract is as secure as its least secure components. If the data supplied by an oracle are corrupted, then so is the output of the smart contract.

The simplest design of an oracle is where an entity queries a single data provider and records the data on the blockchain. For example, it could be a query from a Coinbase web API. This is called a centralized oracle, which is often a fast and efficient solution. However, reliance on one centralized entity and one centralized data source introduces several potential points of failure. First, the entity can be corrupted. For example, the oracle could withhold the data or front run on information it provides. Second, the data can be corrupted in the process of transferring from the data source to the blockchain because of a software bug. Finally, the data source itself needs to be trusted.

In its perpetual quest to minimize trust from relying on third parties, the crypto community has been actively working on new oracle designs. Inspired by the decentralized trust model of permissionless blockchain protocols, decentralized oracles have become one of the fastest-growing solutions, with Chainlink currently dominating the space. The main idea behind any decentralized oracle is to source data from a large and heterogeneous set of entities (nodes) to determine the validity and accuracy of the data and to keep the entities honest by using incentive mechanisms and skin in the game.

Similar to PoS protocols, every participating node that delivers data has to stake a deposit, typically in the native token of the network. If the node provides accurate data, it earns a reward. If it misbehaves, the node can lose a percentage of its stake and, in some cases, access to future participation in the oracle network and, as a result, all future revenue from the protocol.

16. See Beniiche (2020) and Caldarelli and Ellul (2021) for surveys of different oracle types.

The fundamental challenge then is to determine what the truth is. In a blockchain, the correctness of transactions is a property of internal consistency (no double-spending). There can be multiple conflicting versions of the blockchain (forks), but there is always one that is correct, and the goal of validators is to agree on which one. In an oracle network, the situation is more complicated. Depending on the nature of the data in an oracle network, there might not be a true report but only its noisy realizations. Therefore, a typical solution to determine the consensus report is to rely on the wisdom of the crowd and use some form of aggregation across reports, for example, taking the median or mean value.

This reliance on a diversified set of data providers, however, exposes the process to the possibility of an adversarial attack, where an adversary bribes the existing nodes or sets up nodes to produce a corrupt report. Equally problematic could be collusion among oracle nodes. If the gains from collusion become very high, the oracle nodes might not care to lose their current stakes or even all future stakes. As a consequence, the oracle's economic rent should be high enough to ensure that its members are to remain honest.

The research on decentralized oracles is in a fledgling state.[17] There are many open questions. For example, holding the size of oracles network-fixed, what design is the most resilient to the bribery attack? Is it optimal to restrict the size of the network or allow a free entry of nodes? Holding economic rent of an oracle fixed, what is the maximum stake that can be written on the oracle's output?

III. The Current Cryptocurrency Landscape

According to CoinGecko, there were over 10,000 crypto tokens with an aggregate market cap of more than $2 trillion as of February 2022. Several classifications have been proposed for crypto tokens.[18] We have found it useful to parse the universe of crypto tokens into the following large categories, depicted in figure 2.

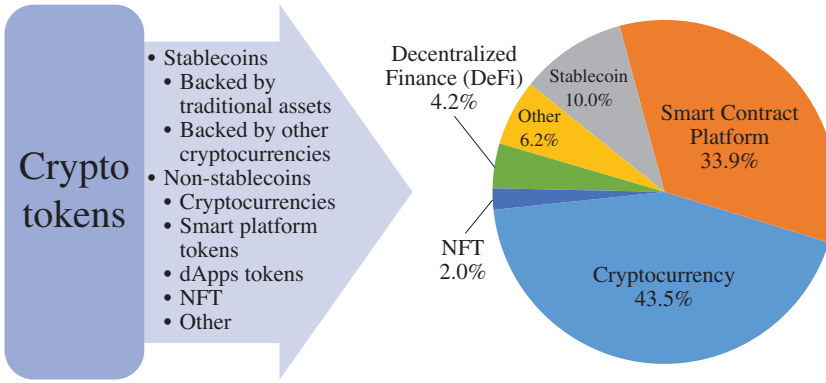
III.A. Stablecoins

To start with, we can separate crypto tokens into stablecoins and non-stablecoins. Stablecoins are designed to maintain a peg to fiat currencies

17. See Breidenbach and others (2021) and the references therein.

18. See, for example, Cong and Xiao (2021) or Prasad (2021). A recent Center for American Progress report describes how cryptocurrencies fit in the current regulatory landscape; see Phillips and Thornton (2022).

Figure 2. Share of Market Capitalization by Token Categories



Sources: CoinGecko and authors' calculations.

Note: This figure shows the share of market capitalization by categories of cryptocurrency tokens and coins (here we collectively refer to them as tokens) as of February 2022. "Smart Contract Platform" includes tokens for platforms that host smart contracts on their own blockchains. "Stablecoin" refers to tokens that are pegged to a specific asset such as fiat currency. The category "dApps" includes tokens used for different decentralized application protocols. "NFT" refers to non-fungible tokens. "Other" refers to the rest of the cryptocurrency tokens that cannot be classified to the categories listed above.

and therefore act as a safe asset that is not subject to the same volatility as many cryptocurrencies. The absence of central bank digital currency and the growth of DeFi applications based on smart contracts created a strong demand for private stablecoins that are native to cryptocurrency protocols. According to CryptoRank, if at the beginning of 2021 the market value of all stablecoins was $30 billion, then by February 2022 it had reached $180 billion.[19] As a point of comparison, the total value of British pound banknotes in circulation in 2021 was about £80 billion.[20]

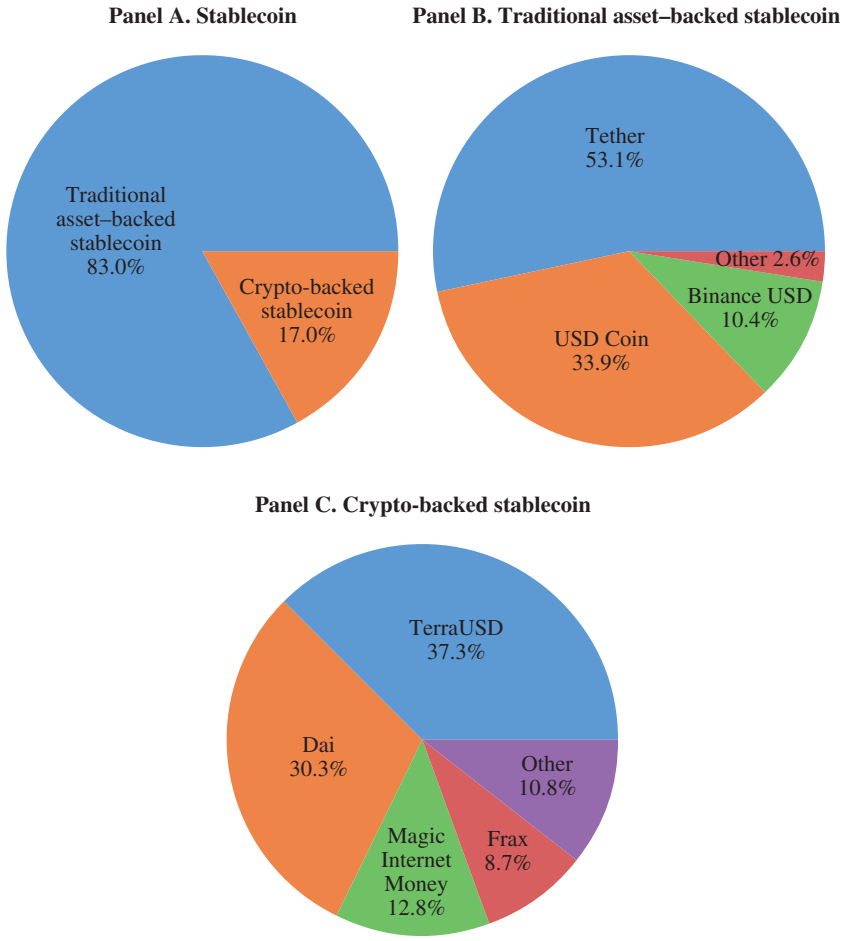
The existing stablecoins can be divided into stablecoins backed by traditional liquid and safe assets, for example, US dollars and Treasury bills, and algorithmic stablecoins backed by other cryptocurrencies. In figure 3, panel A shows the relative share of stablecoins backed by traditional and crypto assets, with the former being the vast majority.

Panels B and C show the largest stablecoins within each category. The stablecoins backed by traditional assets are dominated by just three coins: Tether, USD Coin, and Binance USD. To guarantee the peg, the stablecoins

19. See CryptoRank, "Crypto Market Insights and Analytics," https://cryptorank.io/.

20. See Bank of England, "Banknote Statistics," https://www.bankofengland.co.uk/statistics/banknote.

Figure 3. Share of Market Capitalization by Stablecoin Categories



Sources: CoinGecko and authors' calculations.

Note: This figure shows the share of market capitalization by stablecoin categories as of February 2022. Panel A shows the share of stablecoins backed by traditional assets compared to those backed by crypto assets. Stablecoins backed by crypto assets include those algorithmically backed by a particular cryptocurrency or by multiple tokens such as tokens in a liquidity pool. Panel B shows the share of top stablecoins backed by traditional assets. Panel C shows the share of top stablecoins backed by crypto assets.

backed by traditional assets should be backed one-to-one by cash or cash-like assets such as US Treasuries. Many stablecoin providers had made claims that their tokens were 100 percent backed by liquid assets, only later to reveal that that was not the case. The famous examples include the two most popular stablecoins, Tether and USD Coin.[21] In both cases, some part of collateral was held in securities subject to default risk. In October 2021, Tether was fined $41 million by the Commodity Futures Trading Commission for making misleading claims about being backed one-to-one by the US dollar.[22]

Along with the stablecoins backed by traditional assets, there has also been growing acceptance of algorithmic stablecoins. Based on data from CoinGecko and our calculations, as of February 2022, the combined value of algorithmic stablecoins exceeded $25 billion, with the largest coins being Dai and Terra USD. The rising popularity of algorithmic stablecoins can again be traced to the desire of the crypto community not to rely on centralized parties. Since fiat currencies are issued by governments, the stablecoins backed by traditional assets depend on trust in government. To break from the need to trust the government, algorithmic stablecoins—or, as they are often called, programmable money—use other cryptocurrencies as a collateral and sophisticated algorithms to regulate the stablecoin supply to maintain the peg.[23]

There are now increasing calls for an urgent regulation of the stablecoins. The main concern is that lack of transparency in reporting of the reserves and inadequate collateral can make stablecoins prone to a run. We get back to these issues in section IV.E.

III.B. Non-Stablecoins

Non-stablecoins constitute a large and diverse group. Their value depends on the current investor sentiment and fluctuates widely over time. First, we can isolate coins that have no other function than being a cryptocurrency, either used for transaction purposes or as a store value. This group includes the first generation of cryptocurrencies such as Bitcoin and Litecoin. By construction, these are the cryptocurrencies that are built on non-smart contract platforms. The majority of these cryptocurrencies

21. See, for example, De and Hochstein (2021) and De (2021).

22. See Commodity Futures Trading Commission, "CFTC Orders Tether and Bitfinex to Pay Fines Totaling $42.5 Million," https://www.cftc.gov/PressRoom/PressReleases/8450-21.

23. See, for example, MakerDAO, "The Maker Protocol: MakerDAO's Multi-Collateral Dai (MCD) System," https://makerdao.com/en/whitepaper; and Kereiakes and others (2019).

are based on PoW blockchains. Early on, crypto enthusiasts hoped that these cryptocurrencies could replace government-sponsored currencies as a transaction medium. However, it quickly became clear that this was infeasible because verifying transactions on public PoW ledgers is slow and highly energy-inefficient. Since then, a new narrative for the benefits of these coins has emerged, positioning them as the new "gold"—a digital store of value. Figure A.1 in the online appendix shows that, as of February 2022, Bitcoin dominated this group with a market share of more than 90 percent, followed by Dogecoin. Dogecoin was created in 2013 by two software engineers, Billy Markus and Jackson Palmer, as a parody of a cryptocurrency that was meant to be worthless. It sharply increased in value and became the first meme coin in 2021 following public support by Elon Musk.

SMART CONTRACT PLATFORMS Another large group are tokens issued by smart contract platforms such as Ethereum, Binance Smart Chain, Solana, and Cardano. In many ways, these tokens are similar to the tokens in the first group. In particular, they can also be used to pay for transactions on the platform and are a claim on the platform's economic value. The reason we separate them from the first group is that cryptocurrencies in the first group offer no intrinsic economic value other than the potential for capital appreciation. Therefore, it is unclear what aggregate risk, other than inflation, they are supposed to be tied to.

In contrast, the value of a smart contract platform depends on the scope and the number of applications run on the platform since they affect the number of transactions and the amount of transaction fees, which in turn influence the price of the platform token.[24] Figures A.2 and A.3 in the online appendix show the development of smart contract platforms. The left panel of figure A.2 shows the evolution of the market value of different platforms. The right panel shows platforms' market share. Figure A.3 shows the growth of the total value locked (TVL) on the platforms. TVL is the overall value of crypto assets deposited in applications run on the platform. It has emerged as a main metric for gauging interest in a particular platform or sector of the crypto industry.

Figures A.2 and A.3 show that smart contract platforms grew exceptionally fast in 2021. If at the start of 2021 the total market value of smart

24. This division into two groups is a simplification since even the Bitcoin blockchain can host other protocols, for example, Omni Layer, or help secure other platforms, for example, Rootstock and DeFiChain. However, presently the scope of these applications compared to those built on smart contract platforms is limited.

contract platforms was around $144 billion, at the end of January 2022 it stood at $683 billion, reaching almost $1 trillion in November 2021. Similarly, the combined TVL across all platforms was $18 billion in the beginning of 2021 and grew to about $177 billion by February 2022.

Figures A.2 and A.3 also show that Ethereum still dominates this space. The market share of Ethereum has been relatively stable at about 50 percent. The high fees on the Ethereum platform, however, have led to the growth of other smart platforms and to an increase in share of applications deployed on them. If in the beginning of 2021 Ethereum completely dominated the space, by the end of 2021 its share declined to 66 percent.

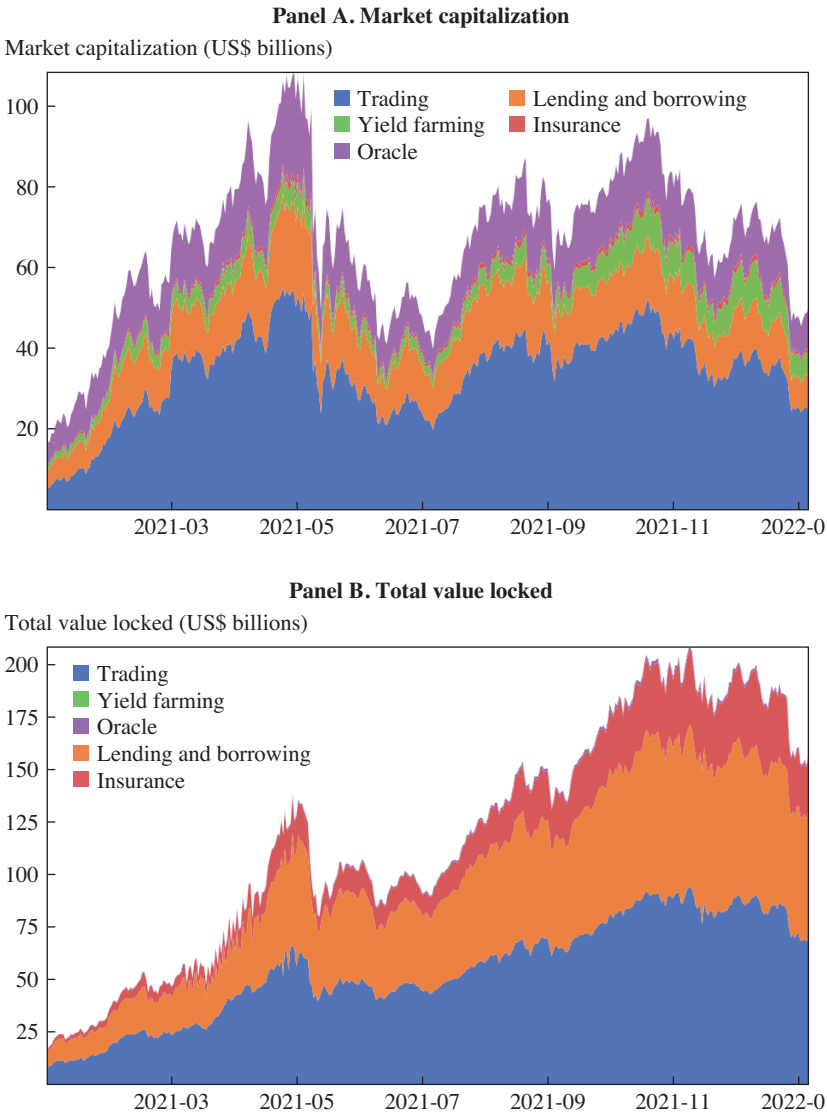
DEFI APPLICATIONS Smart contracts layered on a permissionless blockchain protocol have given rise to the emergence of what is called decentralized finance (DeFi)—a suite of financial applications meant to replicate many of the elements of the traditional financial system without relying on centralized intermediaries.

Figure 4 shows the five largest DeFi sectors. The main applications so far have been centered on trading platforms, lending and borrowing marketplaces, oracles, yield farming, and insurance. Panel A shows the evolution of the market value of the sectors; panel B shows the total value locked in each sector.

Decentralized crypto exchanges. Decentralized exchanges (DEXs) have attracted a lot of attention and have become the fastest-growing sector of the DeFi universe. One of the main advantages of DEXs over centralized exchanges is the ability for users to keep control of their private keys. When market participants deposit their crypto tokens with a centralized exchange, they forfeit their ownership to the exchange. This makes them exposed to exchange risk—if the exchange is hacked and its funds are stolen, investors can experience significant losses. More generally, trading on a centralized exchange requires participants to trust in the exchange, which goes against the maxim of decentralized finance. Trading on DEXs is governed by smart contracts and eliminates counterparty risk for the investors. The settlement of transactions is instantaneous, after they are confirmed and included on the blockchain.

The majority of DEXs use an automated market maker (AMM) protocol, which allows a direct exchange of two crypto tokens, say X and Y. The main object in an AMM protocol is a new market structure called a liquidity pool. A liquidity pool consists of two pools: one of X tokens and one of Y tokens. The ratio of tokens in each pool defines the current exchange rate between the two tokens.

Figure 4. Market Capitalization and Total Value Locked of Decentralized Finance



Sources: CoinGecko, Defi Llama, and authors' calculations.

Note: These figures show the market capitalization and total value locked for different categories of decentralized finance from January 2019 to February 2022. Trading refers to tokens used in decentralized exchanges, including those for spot trade and derivative exchanges. Lending and borrowing refers to DeFi platforms where lenders add funds into liquidity pools in return for a regular interest rate from borrowers. Yield farming includes yield aggregators and protocols that incentivize people to deposit or lend out their tokens in exchange for rewards.

A liquidity pool supports two main operations: liquidity provision and a swap between the two tokens. Anyone who owns the two tokens can choose to be a liquidity provider by depositing tokens X and Y to the respective pools in the proportion equal to the current ratio. In return, the liquidity provider receives a claim on the share of the two pools' tokens, the so-called liquidity pool (LP) tokens.

A swap order allows one to exchange one token for the other. The exchange rate depends on a particular implementation of the AMM protocol and is determined by some deterministic rule called the bonding curve. For example, in the constant product AMM used by a popular DEX, Uniswap V2, if the initial amounts of X and Y tokens in the liquidity pool are x and y, and someone wants to exchange Δx of X tokens for Y tokens, the exchange rate is determined according to the following rule:

$$\left(x + \Delta x\right) \cdot \left(y + \Delta y\right) = x \cdot y \quad \Leftrightarrow \quad \frac{\Delta y}{\Delta x} = -\frac{y}{x + \Delta x}.$$

Swapping X for Y increases the relative share of X tokens in the liquidity pool and therefore lowers its price relative to the price of Y tokens. Whenever the equilibrium price of the two tokens deviates from the current ratio in the two pools, one can profit from it by executing a swap order until the ratio reaches the equilibrium price. To compensate liquidity providers for providing liquidity, everyone who executes a swap order pays a transaction fee that goes to the liquidity pool. This is similar to limit order book exchanges, where liquidity takers executing a market order usually pay liquidity providers who supply limit orders.[25]

The DEX's smart contract usually allows trading any pair of tokens supported by the underlying blockchain. For example, Uniswap V2, realized on the Ethereum blockchain, allows trading any pair of ERC-20 tokens. If no liquidity pool exists for a particular pair of tokens, it can be freely created. The viability of the pool then depends on the ability of the pool to attract liquidity providers and traders. The liquidity is usually concentrated in a few pairs. Figure A.4 in the online appendix shows how DEX trading volume compares against centralized exchanges. While the volume of DEX has experienced fast growth, it still constitutes only a fraction of the centralized exchange volume.

25. See Aoyagi (2020), Aoyagi and Ito (2021), Lehar and Parlour (2021), and Capponi and Jia (2021) for further results and comparison of decentralized and centralized exchanges.

Similar to centralized exchanges, a few DEXs dominate the space. In figure A.5 in the online appendix, the top panel shows the market share of the top ten centralized exchanges, the bottom panel shows the top ten decentralized exchanges. The majority of centralized exchange volume is concentrated on offshore exchanges such as Binance, Huobi, OKX, and FTX, which are subject to little or no regulatory oversight. Similarly, Uniswap, PancakeSwap and SushiSwap account for about 70 percent of volume among DEXs.

Borrowing and lending. Lending protocols have been another fast-growing sector of DeFi. Similar to DEXs, lending and borrowing are governed by smart contracts. The vast majority of DeFi lending is over-collateralized loans secured by other crypto coins, which is primarily used for creating leveraged trading positions.

A typical transaction involves borrowing some of the stablecoins and putting up Ethereum or Bitcoin as a collateral. Since the value of Ethereum or Bitcoin fluctuates, there is a danger that the value of collateral can be lower than the borrowed amount. To mitigate this risk, a smart contract uses an oracle to obtain up-to-date cryptocurrency prices and automatically liquidates the position if the loan-to-value falls below a specified threshold. The threshold depends on the perceived riskiness of the collateral token and ranges between 50 percent and 80 percent.

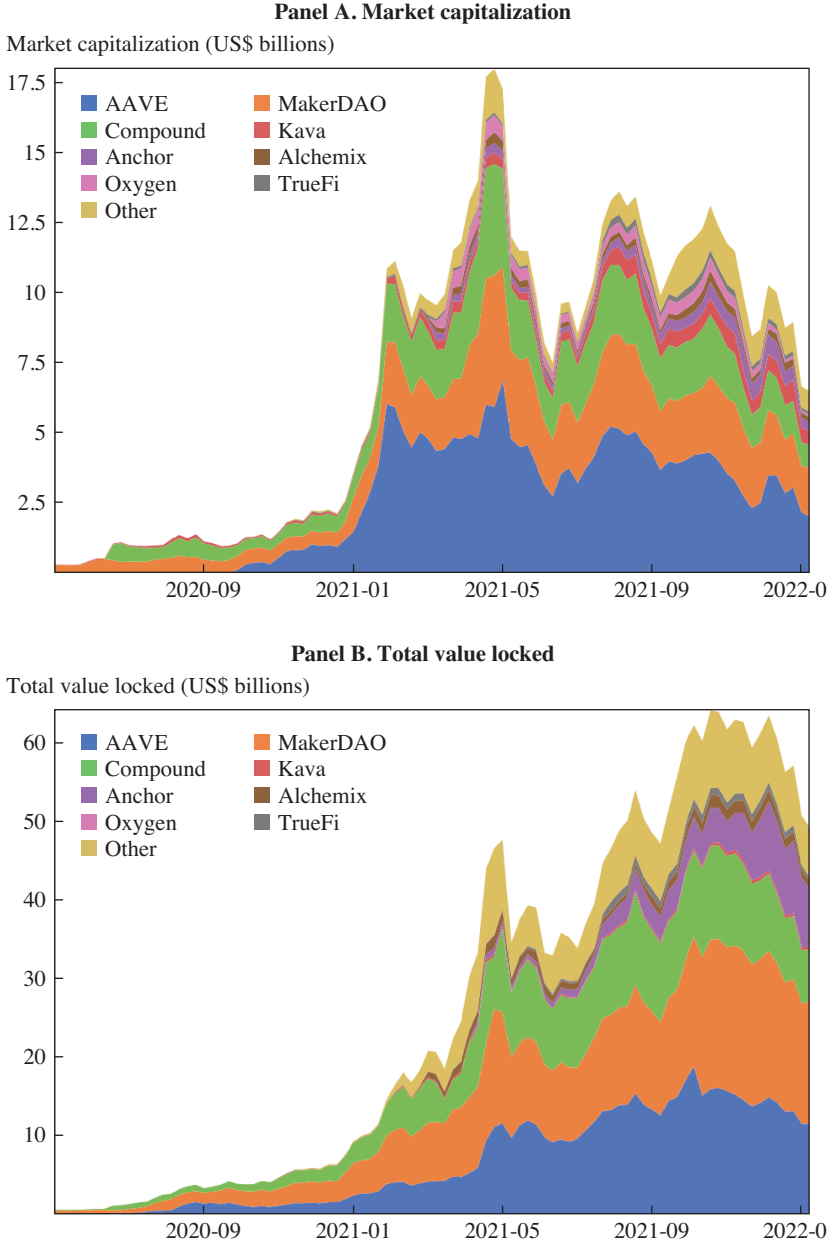
A borrower has to pay a borrowing interest rate and can receive a lending rate on their collateral. In addition, a protocol collects a fee for its service, which goes to the pool controlled by protocol token holders. The lending rate is a function of the borrowing rate and the utilization of funds: borrowing fees, net of protocol fees, are spread among all lenders. The borrowing rate depends on the asset. It is set by the smart contract to maximize utilization of funds and changes in response to the market conditions.

Figure 5 shows that, similar to a DEX, the lending space is dominated by a few large players such as Aave, Anchor, and Compound protocols. Most protocols operate on a few chains; for example, Aave is built on three smart contract platforms: Ethereum, Avalanche, and Polygon. Anchor uses only Terra, and Compound only Ethereum. Thus, the concentration within a particular smart contract platform is even higher.

Figure A.6 in the online appendix shows the aggregated amount deposited and borrowed across different crypto tokens. The main activity is concentrated in stablecoins, along with Ethereum and Wrapped Bitcoin.[26]

26. Wrapped Bitcoin is an Ethereum token that is intended to represent Bitcoin on the Ethereum blockchain. It is backed on a one-to-one basis with Bitcoin.

Figure 5. Market Capitalization and Total Value Locked of Decentralized Lending



Sources: CoinGecko, Defi Llama, and authors' calculations.

Note: This figure shows the market capitalization and total value locked for the top twenty lending protocols based on market capitalization from May 2020 to February 2022.

A large imbalance between the amount deposited and borrowed for Ethereum and Bitcoin means that investors use them as a collateral to borrow stablecoins, which can be used, for example, to buy Ethereum and Bitcoin, thus creating a leveraged position.

Yield farming. The desire to earn supersized returns led to the proliferation of smart contracts that aim to maximize the yield from holding crypto tokens. As we showed above, crypto investors have several strategies to earn return on their coins. First, they can delegate their coins to validators who stake the coins and earn rewards for verification of transactions. Second, investors can earn fees for providing liquidity to DEXs. Third, they can earn interest by depositing their coins into lending protocols. Finally, some token providers use airdrops—the practice of giving away tokens to a subset of investors meeting particular criteria.

The return on any of the above strategies varies over time. Yield farming smart contracts (or simply yield farms) aim to optimize the return by optimally allocating investments among multiple protocols and DeFi applications. The process also usually involves high leverage. For example, LP tokens obtained after placing tokens in a liquidity pool can be further used as collateral or deposited into lending protocols.

The high leverage creates a risk of large losses due to a chain reaction of multiple contracts being liquidated when some contracts lose their value, either during downturn market movements or because of hacks. Also, while yield farm strategies are designed to maximize the yield on investment, they do not automatically result in high returns because the underlying crypto tokens can lose value. In many cases, high yields are financed through an increase in the token supply where the net effect depends on investors' willingness to absorb an ever-increasing supply of tokens.

NFT Lastly, 2021 saw a meteoric rise in hype and value of non-fungible tokens (NFTs). An NFT is a unique piece of data stored on a blockchain. The data can be associated with a particular digital or physical asset or a license to use the asset for a specified purpose. Because each token is uniquely identifiable, NFTs differ from other cryptocurrencies. NFTs can be bought and sold and are seen as a form of digital art. The NFT space attracted attention in March 2021 when a digital collage of 5,000 images by the artist known as Beeple was sold for an eye-popping price of $69 million at Christie's auction house. The combined value of all NFTs at the end of January 2022 stood at about $13 billion.[27]

27. NFTGo, "Market Overview," https://nftgo.io/overview.

IV. DeFi versus the Traditional Financial System

Many of the existing problems with intermediaries originate from well-known economic frictions that are inherent in financial markets, such as asymmetric information, adverse selection, moral hazard, and so on. This creates opportunities for abuse and also significant costs of guarding the public and the economy against financial fraud, malfeasance, and systemic risk. Technological innovations have a long history in finance of helping to provide solutions to the above problems and improving the efficiency of financial markets.

DeFi applications thus far have had limited scope; they have been mainly built around simple applications, such as trading in cryptocurrencies or collateralized lending. But they are growing rapidly in scope and complexity. They have also escaped the burden of regulation and consumer protections and have benefited from tremendous investor optimism that allowed many problems and inefficiencies to go unnoticed.

In what follows, we aim to highlight the important trade-offs offered by the two architectures. When comparing the potential benefits of DeFi solutions with those offered by the traditional system, it is important to think about the proposed new solutions in the context of the larger financial architecture rather than narrowly focusing on individual dimensions of possible inefficiencies.

IV.A. Data Privacy and Transparency

How to protect data privacy in an increasingly digital society has become a major concern to regulators, activists, and regular citizens alike. Crypto enthusiasts often tout the anonymity of transactions as "a feature, not a bug" and view it as a major benefit over the traditional model, where the failure or corruption of a centralized intermediary could lead client data to be mistakenly exposed or hacked. While it is in the commercial interest of intermediaries to protect the privacy of their clients, it is a reasonable concern that intermediaries might not endogenize the full cost to the clients.[28] This conflict leads to a classic underinvestment problem relative to what consumers would prefer. In addition, financial intermediaries might have an interest in using client data for their own commercial purposes or allowing third parties access, including the government.

28. Some infamous recent examples of data breaches in the financial sector are the 2017 breach of Equifax that exposed personal information of 147 million people and occurrences at banks like Capital One and First American Financial Corporation; see Tunggal (2022).

Recognizing this problem, in the United States a large set of regulations, such as the Bank Secrecy Act, Right to Financial Privacy Act, the Gramm-Leach-Bliley Act, and the Fair Credit Reporting Act, has been put in place to protect consumers from unlawful access to their financial accounts by private and public institutions and the unlawful disclosure or commercial use of financial information.

But the laws also recognize an important trade-off between individual privacy and other important societal goals, such as preventing malevolent actors from using the financial system for money laundering, financing of criminal and terrorist activities, or tax evasion. This is typically achieved by putting into place know your customer (KYC) and anti–money laundering (AML) laws that require financial institutions to verify the identity of a client when opening an account and to provide government authorities with information about suspicious financial transactions. Financial intermediaries in the traditional system then play the dual role of acting on the one hand as a shield to prevent the unauthorized collection, use, and disclosure of sensitive data. But on the other hand, they selectively grant access to information in well-defined circumstances where access to such data is important for the functioning of the economy or the broader society. Examples include reporting of capital gains tax to the IRS or granting access to financial accounts of individuals in cases where an illegal or terrorist intent has been clearly defined by law and regulation.

Cryptocurrencies built on permissionless protocols preserve privacy by design by not collecting any personal information about account holders. Crypto tokens are represented by alphanumeric strings and protected by cryptography algorithms. Crypto addresses are very easy to generate, and many protocols encourage users not to use an address more than once. Even if a protocol has a complete record of transactions, the identity of the person behind the transactions cannot be established unless this person uses the tokens to transact with an entity that does enforce KYC norms, such as a regulated financial institution. In many ways, the current modus operandi of cryptocurrencies is similar to an old Swiss model of banking where people could set up anonymous accounts and no questions were asked. This model, however, has been rejected in the majority of developed countries in favor of more transparency and accountability.

Collecting and protecting data is not costless, and in the traditional architecture, intermediaries bear this cost. The benefits of relying on intermediaries as the important entry nodes for participants in the traditional financial system mean that KYC norms or AML laws have to be monitored only at a limited set of nodes. For example, when a customer makes

a payment using a credit card or a bank transfer from a US bank, a retailer does not need to worry about the legality of the funds. Similarly, the ability to collect taxes depends on the government's capacity to trace transactions and link them back to a person or organization. In the traditional system, centralized intermediaries such as exchanges or brokers are responsible for reporting transactions to the IRS.

The permissionless and pseudonymous architecture of DeFi generates formidable challenges for tax enforcement, aggravates issues of money laundering and other kinds of financial malfeasance, and as a result creates externalities on the rest of the economy. If entry into the system is not monitored by intermediaries but happens completely anonymously by setting up an address on a blockchain, KYC norms and AML laws would need to be regulated at the level of the transaction. In many cases this could be prohibitively costly or impractical and therefore lead to an untransparent environment that facilitates illegal transactions.

Consider, for example, trading on a DEX. Recall that a DEX is simply a smart contract that executes trading between any pair of cryptocurrencies and that can be deployed anonymously by anyone. Suppose a customer trades and realizes some capital gains. Since the identity of the person behind the transactions cannot be established until this person uses the tokens at an entity that does an identification check, by transacting with entities that do not verify identification, the person could spend the tokens linked to the capital gains transactions and thus avoid ever paying capital gains taxes.

But even if the person transacts with an entity that does enforce KYC standards, this does not reveal any capital gains associated with the past transactions of this coin. In order to impute the true capital gains tax, the entity would need either to investigate the full history of transactions up to the current point or to delegate this task to another intermediary. In practice, tracing transactions along often multiple protocols is a challenging problem. Specialized blockchain analytics companies such as Bitfury Crystal and Chainalysis have shown that it can be done successfully in select cases of illegal transactions. However, successfully tracing all transactions would likely be very costly. Makarov and Schoar (2021) show, for example, that Bitcoin flowing out of dark net markets like Hydra can be laundered through many intermediary addresses and can eventually enter KYC-compliant exchanges such as Coinbase or Gemini without being tagged.

The pseudonymous nature of cryptocurrencies also makes it much harder to enforce rules against market manipulation, insider trading, and

self-dealing, since suspicious transactions cannot easily be traced back to individuals. For example, large holders of cryptocurrencies have strong incentives to lobby government officials or regulators to promote investments in cryptocurrencies and adopt lax regulation. Especially at the early stages in the development of new technologies, any announcements endorsing the official use of cryptocurrencies create significant positive price impact (Auer and Claessens 2020). The danger is that some regulators or politicians (or their friends) receive gifts in the form of cryptocurrencies (or simply already own cryptocurrencies) which would tilt their decision toward adoption even if it is not in the interest of the general public.

As the above discussion shows, to safeguard society against these inherent risks, a completely new framework of ensuring KYC and AML standards would have to be developed. The majority of DeFi players actively lobby that they should not be bearing the costs of linking transactions to economic actors and ensuring that the financial system preserves an adequate level of transparency and accountability, citing technological constraints or the danger of losing a competitive advantage in the crypto space.[29] But unless society gives up entirely on collecting taxes and implementing KYC and AML practices, somebody has to bear these costs.

IV.B. Economic Rents

Another important dimension by which to assess a financial system is how economic rents are distributed among agents in the system. An important concern with the traditional financial system has been that the centralized position of intermediaries can allow them to extract excess economic rents at the expense of their customers. The proponents of the DeFi architecture typically argue that the open-source and permissionless nature of DeFi protocols promotes competition. Therefore, the claim is that DeFi solutions should drive out excess rents.

This view, however, neglects the fact that free entry is not synonymous with more competition and thus not a panacea for beneficial outcomes in many situations. The effectiveness of competition depends on a number of factors, such as whether there are barriers to entry, switching costs, product differentiation, asymmetric information, and network externalities.

29. For example, see Staking Facilities, "Staking Infrastructure Providers Unite in the European Blockchain Association," https://stakingfac.medium.com/staking-infrastructure-providers-unite-in-the-european-blockchain-association-6ecebbb8139f; Financial Services Republicans, "McHenry Leads Bipartisan Letter Urging Yellen to Clarify Digital Asset Reporting Requirements," https://republicans-financialservices.house.gov/news/documentsingle.aspx?DocumentID=408238.

The presence of any of these factors hinders competition, and in some cases even creates adverse effects from competition. Technological changes that affect any of these factors, therefore, also transform the competitive landscape.

Similar to the traditional financial system, there are several natural points where rents can accumulate at different layers in the DeFi architecture due to endogenous constraints to competition.

First, at the level of validators of transactions, in both PoW and PoS rents can accumulate due to inherent economies of scale and scope. In theory, in PoW protocols, if miners were fully decentralized, one could expect them to earn zero rent in a steady state because of free entry. In practice, however, as we showed in section I.A, mining is concentrated in pools and at the level of individual miners. High concentration of mining power can facilitate collusion and help sustain transaction fees above their average costs. For a dominant protocol such as Bitcoin, the competition from other PoW protocols can be limited because of the negative externalities the dominant network has on the security of smaller PoW networks. In particular, mining capacity can be redirected to launch 51% attacks on the smaller networks, as discussed in section I.A.

Similarly, rents can also accrue to validators in PoS protocols. We showed in section I.B that validators in PoS are concentrated. Furthermore, the same validators are active over a large cross-section of cryptocurrencies, effectively forming a new market structure. These validators control a large proportion of wealth that gives them substantial competitive advantage over newcomers with small amount of wealth.

Second, rents can also accrue at the level of the smart contract platforms that are built on the base layers. Similar to traditional payment systems like Visa, Mastercard, or PayPal, there are strong network externalities. Smart contract platforms differentiate themselves by the choice of programming language to code up smart contracts and the network architecture and often have a limited degree of interoperability. While smart contracts built on the same protocol can interact seamlessly with each other, communication between applications built on different platforms in general is limited.[30]

Naturally, the decision of which platform to build an application on depends on the existing pool of applications already deployed on the platform and the platform's future growth prospects. A popular platform with

30. A number of solutions have been proposed and are being developed to increase interoperability between chains; see, for example, Ethereum, "Blockchain Bridges," https://ethereum.org/en/bridges/, for more details.

a wide range of applications and a large user base provides better business prospects and therefore is more attractive than a less popular platform. Often these network effects increase exponentially with each user. As a result, developers and users might choose a more popular platform even if it charges higher transaction fees. These network externalities might also stand in the way of switching to a platform with a better technology if a critical mass of users is captured by the incumbent platform.

One could argue that even if the platform is a monopolist, competition between validators on that platform will keep fees low. However, as we showed above, high concentration of validators can lead to collusion and allow them to earn excess rents. Even if validators do not collude, high transaction fees can still be realized if the platform operating capacity is limited and users need to pay a premium for priority execution (Huberman, Leshno, and Moallemi 2021). Finally, the majority of PoS protocols have a minimum level of transaction fees as a protocol parameter, which provides the platform with a direct tool to limit competition among validators and earn rent.

Figure A.7 in the online appendix shows total transaction fees in the year 2021 across different platforms. The case of Ethereum is striking. The platform generated nearly $10 billion in fees from about 460 million transactions. In contrast, Visa's total revenue was around $24 billion over 164.7 billion transactions.[31] Thus, an average Ethereum fee per transaction has been one hundred times that of Visa.

For PoS platforms, an added complexity arises from the fact that the going concern value of the platform also affects the security of the platform itself and the applications that run on it. Since the value of the platform depends on the level of transaction fees, fees should be high enough to deter possible attacks on the platform, which can further support the platform's rent in equilibrium. These security concerns can also decrease competition among platforms. Since a low-value platform can be more easily attacked, the concerns over the platform's security may lead to slower growth, which in turn can reduce the platform's current value.

Third, economies of scale at the level of individual DeFi applications can allow them to assemble local monopoly power and extract rents despite the open-source architecture of the blockchain. In addition, while in theory crypto smart contracts are usually described as open-source code, in practice successful applications have tried to protect their code

31. This figure is larger than transaction fees alone since Visa earns revenue from sources other than fees paid by direct users.

and limit its distribution. Here, an example of two DEXs, Uniswap and SushiSwap, is instructive.

Originally, Uniswap V2 was operated as open-source software utilizing a general public license, which allows anyone to run, distribute, or modify its code. This has been used by a pseudonymous developer called Chef Nomi to create a clone of Uniswap called SushiSwap. Similar to centralized exchanges, DEXs are subject to economies of scale. An exchange with a large liquidity pool is preferred over an exchange with a small one. Therefore, an exchange clone will typically find it difficult to challenge the original exchange.

To compete with Uniswap, SushiSwap introduced a new business model, which has now been adopted by a majority of other applications. The main change made by Chef Nomi was to create a governance token (SUSHI) and give it as a reward to traders who provide liquidity to the platform. The token allows its holders to vote on how the SushiSwap platform is run and potentially receive a portion of the transaction fees. As a consequence, investors can trade these tokens and speculate on the future prospects of the platform. This business model strengthens network externalities and therefore limits copycat strategies and competition. The more valuable the platform and its tokens are, the higher is the reward for liquidity providers. A larger liquidity pool, in turn, attracts more trading on the platform, which makes the platform more valuable.

The SUSHI token was also used to launch a "vampire attack" to drain liquidity out of Uniswap, whereby SUSHI tokens could be exchanged for Uniswap LP tokens. Those LP tokens would then be exchanged for the original assets put into the Uniswap liquidity pools, thus creating liquidity for SushiSwap instead. The attack was successful, draining Uniswap of about 55 percent of its liquidity (Gushue 2021).

In response, Uniswap introduced its own governance token (UNI). To limit copycat attacks, the new version of the protocol, Uniswap V3, also adopted a different license agreement, called business source license, which incorporates copyright law and allows Uniswap governance to restrict unauthorized commercialization of an entity's source code for two years.

Finally, rent extraction can be driven by frictions at the customer level due to lack of financial literacy or behavioral biases. Many financial products today, including smart contracts, are complex contracts with multiple features. If consumers lack the financial sophistication to understand these product features, institutions that issue these contracts can shroud the actual cost of a product or service. A typical shrouding technique is to advertise or draw attention to one set of attractive features but hide other

more expensive ones. If consumers are unable to analyze what is the best product, even competition might not prevent rent extraction. In fact, more competition might lead to more shrouding as competing firms try to appeal to consumers with evermore enticing and salient features while hiding the unappealing dimensions of the product. Consumer finance products are often designed and marketed in this fashion, which leads to differential targeting of customers based on their financial literacy.[32] Similarly, in the crypto space, practices such as airdrops, yield farming, and meme DeFi tokens have helped capture interest of many investors, but many industry insiders question their value (Di Salvo 2020; Stevens 2020).

IV.C. Transaction Costs

Even if a financial system limits economic rents, it can still be inefficient because of high transaction costs. The traditional financial system has many inefficiencies, which result in high costs of banking services and long settlement time of transactions. A substantial part of these costs comes from the need to cover brick-and-mortar costs of traditional banks and outdated infrastructure. Many banks today still use customized software from the 1980s that lacks real-time account reconciliation and liquidity management capabilities.

While many technological advances are largely exogenous to banks' actions, the decision when and how to implement them depends on the financial architecture. Centralized intermediaries can have limited incentives to invest in new technologies that could threaten their centralized position even if they are welfare improving. Also, modernizing a bank's internal system can have a limited effect if other banks do not coordinate on the change. Often the threat of losing business to new entrants is necessary to force the incumbents to adopt more efficient technology.

The development of blockchain technology has certainly had a positive effect on incentives for the financial industry to upgrade its infrastructure and reduce costs. It is less clear, however, to what extent the potential to reduce the costs depends on the permissionless nature of blockchain. In many cases, arguments can be made that a permissioned blockchain could be designed to deliver a more cost-efficient and robust solution without curtailing competition.[33]

32. See, for example, Célérier and Vallée (2017) and Ru and Schoar (2016).

33. See, for example, SWIFT, "SWIFT Completes Landmark DLT Proof of Concept," https://www.swift.com/news-events/news/swift-completes-landmark-dlt-proof-concept.

Notice also that the permissionless and open-source nature of a protocol does not necessarily make an innovation process easy. It is often argued that if a blockchain protocol is inefficient, then one can create an improved version (aka hard fork) by copying and upgrading the existing code. We showed in section IV.B that competition can be limited between different protocols because of strong network externalities and miners or validators can earn rent in equilibrium. If a new fork leaves less rent to miners and validators, they can have limited incentives to support it. Bier (2021) details the fight among Bitcoin developers about the Bitcoin protocol parameters that occurred in 2015–2017 and provides additional insights into challenges that come with forking a competing blockchain.

IV.D. Governance

The promoters of cryptocurrencies often highlight the idea that the blockchain ledger removes the need for a trusted third party in the execution of contracts. However, this does not mean that the system can function completely devoid of any human intervention. Even if the execution of transactions and smart contracts on the blockchain are automated, the rules governing the blockchain itself and any upgrades to the system must be agreed upon and implemented by its participants. These rules define the governance of the system and in turn how it represents the interest of its different stakeholders.

The major stakeholders in a blockchain ecosystem are, first, the core developers who are charged with writing and updating the code that runs the blockchain. The validators who verify transactions and ensure the integrity of the blockchain are the second set of stakeholders. Often, they decide if they want to adopt the changes provided by the developers. The third important group are the token holders. We can think of these as investors or equity holders. Finally, the fourth group are users of the platform. On some platforms, the third and the fourth groups are the same people.

While all stakeholders have an interest in making the cryptocurrency they are engaged with succeed and grow, their incentives are not always completely aligned. For example, the users and developers might want fees on the blockchain to be low to make utilization more attractive, while investors and validators want to maximize the return on their financial investments. Stakeholders might also differ in their nonpecuniary benefits; for example, some participants might be willing to forgo economic benefits for other objectives, such as maintaining the independence or purity of the blockchain or possibly to undermine other blockchains, as discussed before.

Thus, the classic problems in governance apply also to the crypto universe: rules have to be set to facilitate coordination and provide incentives to adopt value increasing investments and to prevent minority stakeholders from being expropriated by powerful insiders. Providers of capital are particularly prone to expropriation, since once the investment is made, they do not have continued value added or recourse to the firm.

Corporate governance has been a prominent issue probably as long as organizations have existed; in academic research the topic has attracted an enormous body of research at least since the publication of Berle and Means's famous book in 1932.[34] While there is significant heterogeneity across countries in specific corporate governance rules, academic research has shown that private solutions even in competitive financial markets cannot generally resolve governance issues, and the recourse to the legal system is a crucial prerequisite for a well-functioning financial system.[35]

But this reliance on legal enforcement clashes with the maxim of DeFi that tries to avoid placing trust in any actor or institution, including the legal ones. In response to this challenge, DeFi has tried to develop a new form of governance, so-called decentralized autonomous organization (DAO). The basic idea of DAO is to spread control over decisions among all interested stakeholders. This is done by issuing special governance tokens that give their holders the power to propose changes to the protocol and vote on them. All activity is governed by smart contracts and recorded on the blockchain. In most DeFi applications, one governance token equals a vote, and new proposals are implemented according to a predefined majority rule. To ensure that the holders of governance tokens have an interest in the success of the platform long term, protocols often channel a share of the network's transaction fees into the wallets of the governance token holders. The tokens may also carry non-governance rights, like the right to be exchanged for certain other tokens at predefined rates. A famous example of DAO is MakerDAO.[36] Here is how DAO is explained on the Ethereum website: "Starting an organization with someone that involves funding and money requires a lot of trust in the people you're working with. But it's hard to trust someone you've only ever interacted with on the internet. With DAOs you don't need to trust anyone else in the group, just the DAO's code, which is 100% transparent and verifiable by anyone."[37]

34. For an overview, see Hermalin and Weisbach (2017).
35. See, for example, La Porta and others (2000).
36. MakerDAO, "MKR Governance," https://makerdao.com/en/governance.
37. Ethereum, "Decentralized Autonomous Organizations (DAOs)," https://ethereum.org/en/dao/.

But while a transparent and verifiable governance process is certainly an important first step, it does not necessarily ensure good governance. Any DAO design faces the same fundamental trade-offs and issues as traditional organizations. First, decision making in a fully decentralized organization can be inefficient. When the ownership is dispersed and stakes are small, no owner might find it in their interest to spend effort and invest in learning about all the complexities needed to make a decision. As a result, many stakeholders might refrain from voting or lend their votes to a party that is trying to amass voting rights for self-interested reasons. Second, there is always a danger that investors with large stakes (blockholders) can capture control and impose their preference on the system. Recognizing this problem, corporate laws usually impose strict disclosure rules on blockholders. Emulating similar rules on a public permissionless blockchain would be challenging since everyone can control multiple anonymous accounts. Third, the voting system can give more power to participants who may only be interested in maximizing short-term profits as opposed to developing the protocol toward innovative use cases.[38] These arguments are very similar to the debate about investor short-termism in traditional governance (Roe 2018).

Not surprisingly, the crypto space is abundant with colorful examples of governance issues.[39] Ultimately, the majority of insiders recognizes the inherent tensions posed by greater decentralization. Figure 6 shows that in the majority of crypto projects, developers and early investors chose to keep control of the platform by allocating significant stakes to themselves. In addition, even if developers do not have a large stake, in many cases they managed to maintain de facto significant control over the platform, for example, Vitalik Buterin, who has been dubbed the "benevolent dictator for life" (Van Wirdum 2016, quoting Charles Hoskinson, par. 19).[40]

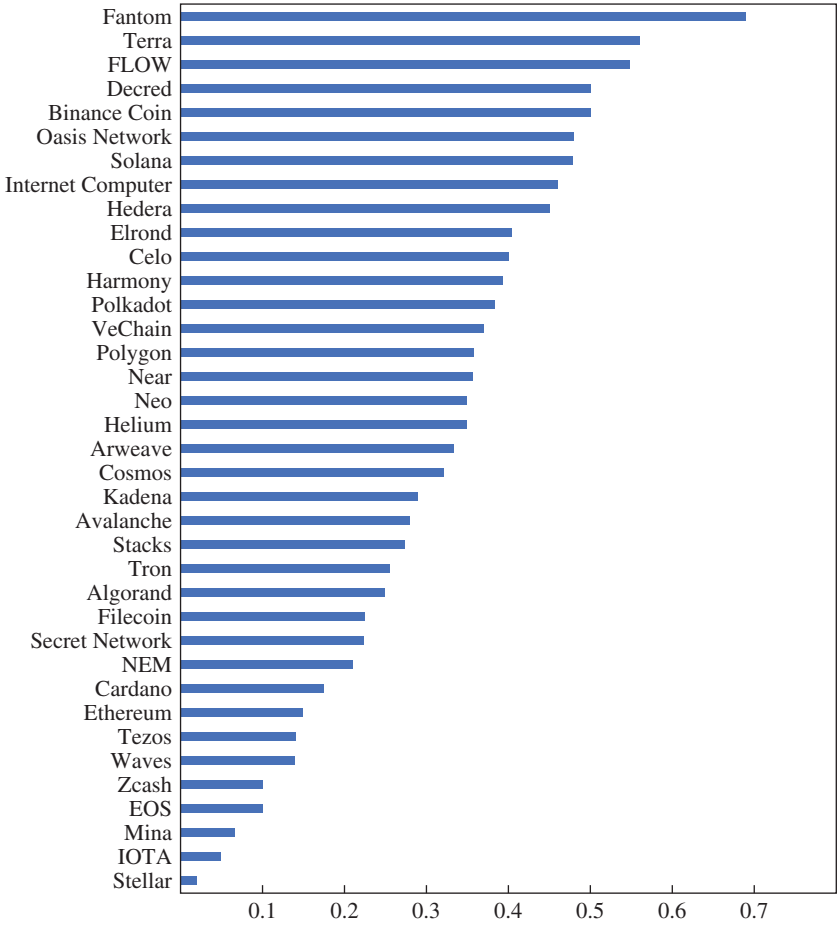
There has been little evidence so far to suggest that the crypto space can successfully resolve governance issues without relying on some off-chain mechanisms. Given that governance issues of blockchain platforms and traditional financial firms are not materially different, it is very likely that robust governance mechanisms will require the support of external regulation.

38. In fact, many recent attacks on DeFi apps exploited the possibility of taking over the voting mechanism to divert funds to the attacker; see, for example, Vigna (2022).

39. See, for example, Bier (2021) and an attempted hostile takeover of Steem (Copeland 2020).

40. Buterin has also been one of the prominent critics of the DAO; see, for example, Buterin (2021).

Figure 6. Initial Coin Offering Insider Share



Sources: Messari.io and authors' calculations.

Note: This figure shows the insider shares from top fifty tokens' initial coin offerings. Insider share includes tokens to founding teams and developers; early investors such as seed investors, venture capital firms, and private sale investors; and associated entities which include companies that are related to the protocols or protocol founders. Insider shares do not include shares that go into the community, such as airdrops, grants, rewards, and tokens to public sale investors, or shares for the development of protocols, such as those going into foundations and reserves.

The history of corporate governance demonstrates that simply providing incentives for managers or investors might not be sufficient to deter bad actors if the financial gains from misbehavior are large. As the implementation of governance rules in the United States has shown, personal accountability of managers and directors is centrally important (Bartlett and Talley 2017). Fiduciary duties that hold corporate agents personally accountable play a critical role in the enforcement of governance rules. The idea is that the threat of punishment creates disincentives for fraudulent behavior, where just losing some money from fraud would not have the same effect.

The pseudonymous nature of the permissionless blockchain environment, however, can make it difficult to hold bad actors accountable for their actions in the same way.

IV.E. Systemic Risk

One of the main sources of systemic risk in the traditional financial system is the reliance on fractional reserve banking. When banks take deposits from the public, they only need to hold a fraction of these deposits in liquid assets as a reserve and can lend the remainder out to borrowers. The goal of a fractional reserve system is to expand economic activities in the economy by freeing capital for lending. It permits banks to use the majority of the deposits to generate returns in the form of interest rates on loans. The efficiency, however, comes at a cost of possible bank failures and runs.

DeFi so far has been operating under a narrow banking model, where every loan is over-collateralized.[41] Narrow banking removes many of the problems faced by fractional reserve systems, but it also constrains the efficient use of capital. The main risk comes from the ability of investors to take leveraged positions and a potential run on stablecoins.

A run on stablecoins can occur for a number of reasons. For stablecoins backed by traditional assets, a run can happen for similar reasons to a run on bank or money market funds. In the absence of timely information about reserves, if investors doubt the quality of the collateral, they have an incentive to exchange the stablecoin for cash, causing a run unless the

41. There have been isolated examples of undercollateralized loans. See Medium, "The Current State of Undercollateralized DeFi Lending—2021," https://medium.com/coinmonks/the-current-state-of-undercollateralized-defi-lending-2021-1f84e14527b5, for an overview of the current solutions.

stablecoin is backed one-to-one with liquid assets like cash or short-term Treasuries. Possible solutions range from issuing stablecoins through insured banks, requiring stablecoins to be backed one-to-one with safe assets, to establishing a central bank digital currency. For a detailed discussion of the regulation of stablecoins and the trade-offs between private and central bank digital currencies, see Gorton and Zhang (2021) and Gorton (2021).

The situation is more complex in the case of algorithmic stablecoins that rely on intricate designs meant to help maintain the peg. Here the issue is less about transparency or misreporting because the design of a stablecoin is public knowledge and all transactions are recorded on the blockchain. Rather, the main concern is about the complexity and potential fragility of the system. Since algorithmic stablecoins are not fully backed by safe assets, it is reasonable to expect that, irrespective of a particular design, there always exist states of the world where the peg is broken and there can be a run on the stablecoin. The available documentation of stablecoins universally lacks rigorous analysis and contains only claims that the design is robust, which, as the case of Iron Finance's Titanium proves, can just be wishful thinking (Lim 2021).

The ability to establish highly leveraged positions is another source of systemic risk. The crypto ecosystem is famous for its wide range of highly leveraged products, with many exchanges offering up to one hundred times leverage for perpetual derivative contracts.[42] Figure A.8 in the online appendix shows that starting July 2021, volume in crypto derivatives exceeded the volume in spot markets. High leverage exacerbates volatility and, as many industry observers believe, is responsible for strong de-leveraging cycles and associated sharp drops in the cryptocurrency prices (McFarlane 2021; Leclair and Rule 2021).

DeFi adds an additional complication to the picture. Many DeFi protocols facilitate leverage and accept other protocols' tokens as collateral. Even though every smart contract and transaction is recorded on a blockchain, and therefore in theory could be analyzed, in practice multiple interconnecting contracts interacting with pseudonymous accounts result in a highly complex and potentially fragile system. This fragility could potentially be exacerbated if some critical smart contracts have unintended coding bugs.

42. See, for example, Potter (2022).

V. Regulation

As discussed, the new financial architecture proposed by cryptocurrencies and DeFi presents formidable challenges for regulators. Regulation of financial assets and services typically has three broad goals: (1) prevent the use of funds for illicit activities, money laundering, or tax evasion; (2) protect participants in financial markets against fraud and abuses; and (3) ensure the integrity of markets and payment systems and overall financial stability.

Our discussion in section IV highlights that at present DeFi solutions do not comply with these three goals. If society does not want to give up on ensuring these goals, some form of technological and regulatory solution seems desirable. In the response to the rise of cryptocurrencies, different countries have followed vastly different approaches. For example, China officially banned trading in cryptocurrencies and developed its own central bank digital currency, while El Salvador allowed Bitcoin as legal tender. In the United States the regulatory environment is still in flux, and there are overlapping responsibilities and sometimes even contradictory approaches.

However, there is urgency to providing a clearer regulatory framework for at least two reasons. First, regulatory certainty is always important for entrepreneurs and investors who wish to decide whether and how to participate in new technologies. Second, the exponential growth of cryptocurrencies can lead to a situation where the political economy of regulation becomes very difficult if regulators wait too long. In effect, cryptocurrencies and DeFi applications can become too big to regulate. We showed in section IV that currently many DeFi solutions do not bear the full cost of the externalities they impose on the economy, such as enforcing KYC or AML laws or complying with tax reporting. Part of the current valuation of some cryptocurrencies and DeFi applications might even be based on an expectation that they will not have to ever comply with these regulations. Thus, requiring DeFi solutions to start internalizing these costs will likely result in losses for some of the current investors. As a result, any such proposals are usually met with strong resistance by the crypto community. This puts regulators in a difficult position. While they need to safeguard the financial system, in a democracy there is often populist pressure to forgo doing the things that are good in the long run to satisfy short-run goals. In fact, the losses might be blamed on the regulatory action itself, rather than the attempt by regulators to prevent even larger losses for society going forward.

The main challenges in regulating cryptocurrencies stem from the pseudonymous and jurisdiction-free nature of this new architecture, which is a consequence of the use of permissionless blockchain protocols and the smart contracts running on them. The traditional financial architecture, where access runs through centralized intermediaries. allows each country to determine its own regulatory framework and decide, for example, who can open a bank account, what documentation must be supplied, and how information can be collected and stored. Also, as the 2022 geopolitical situation between Russia and the West shows, the traditional system makes it possible to restrict the financial system of one country from accessing the financial systems of other countries.

The anonymous and permissionless nature of DeFi apps and the underlying blockchain protocols have the potential to remove the boundaries between the financial systems of different countries or even enable citizens to transact in an ecosystem that is completely outside of government regulation or tax enforcement. While financial integration can have benefits through better risk sharing or improved liquidity, it can also have large costs if poorly regulated systems undercut better regulated ones in a race to the bottom. This becomes especially prevalent if different financial systems operate with vastly different standards.

So what are the available options for regulators? While a complete discussion of all dimensions of regulation is beyond the scope of this paper, we outline a few key options for rule makers. A natural place for regulatory oversight in this new ecosystem is at the level of developers and validators, who in turn control the network protocol. Once this level of regulatory compliance is established, many other functions can be built. In particular, separate entities can be established that would be responsible for verifying the identities and certifying that crypto addresses belong to confirmed users. These entities should be subject to regular audits. The protocols can be adjusted so that validators can check if a particular address belongs to a certified entity, and validators would be charged with only processing transactions that involve certified addresses.

In addition, one could imagine that customers can also be provided with private keys based on their characteristics, such as financial wealth or sophistication. Smart contracts can be ranked based on their safety, risk, and so on. Rules can be established that would allow different smart contract categories to interact with customers who can provide the required key. Smart contracts can be designed to automate the ranking of other smart contracts and automate the generation of private keys. Cryptography algorithms can be developed to guard customers' privacy. Transitioning to

this model will likely require some time and development of new solutions. Therefore, it would be important to lay out an appropriate timeline and deadlines so that market participants can prepare for a smooth transition.

Since countries might differ in how they want to structure their regulatory environment for validators, each country can opt to run its own version of the blockchain. But if some countries agree broadly on regulatory standards, they can use the same blockchain. Countries that choose to run separate versions of the blockchain can interact with others using interoperability mechanisms such as bridges. The above solution can be more easily applied to new blockchains. But if a majority of large countries agree on coordinated regulation, then even the existing blockchains can be brought into a legal framework without the need to break them up into separate sidechains based on different regulatory requirements.

The above solution looks similar to a permissioned blockchain, but this system preserves most of the desired properties of the original design of cryptocurrencies; for example, transactions can be observable on the blockchain, settlement is immediate, and the same set of smart contracts can be executed on it. In addition, if many countries agree on regulation, validators can be elected so that no country has a monopoly over the networks. The ability to regulate validators can potentially change the enforcement of smart contracts by allowing recourse to the contracting parties. But, as we discussed in section II, it can have a positive effect on efficiency.

In contrast, if regulators give up on the ability to oversee validators, the effectiveness of regulation will be much more limited and will depend on the goodwill and voluntary cooperation of validators and developers of the blockchain. If validators accept transactions from every party, the most regulators can hope for is to separate the network into regulated and unregulated parts. This could be done, say, by requiring US citizens to interact only with certified DeFi apps which comply with KYC and AML regulations and provide reports on trades, tax compliance, or other activities. The relative size of the regulated and unregulated networks will depend on the relative investment opportunities in these two networks and the ease of moving funds between them. The problem of regulating compliance only at the level of DeFi apps is, first, that many citizens even from countries that try to regulate DeFi apps could still find it attractive to invest funds in the unregulated network to avoid paying taxes and the like. The ability to evade compliance can provide a large subsidy for the unregulated part

of DeFi apps. Second, regulation will have generally a limited bite on the unregulated part, which can harbor many bad actors and facilitate illegal activities. The opportunities to sidestep the regulated part will generally increase with the level of crypto adoption, since people will be able to interact predominantly in the unregulated part and avoid triggering regulatory compliance.

VI. Conclusion

In this paper we provided an introduction to how the new DeFi architecture works and the mechanics behind it. We also laid out some of the potential benefits and challenges of the developing new system and presented a comparison to the traditional system of financial intermediation. In our discussion we focused on the economic forces and frictions that can arise within this system and the regulatory approaches that might help to mitigate the problems. Our analysis highlights that while the DeFi architecture might have the potential to reduce transaction costs, it is not an automatic solution to the problem of rents in the financial sector. And it may also create additional problems. We identify as a key challenge to regulators the permissionless and anonymous nature of the current DeFi blockchains. These provide the opportunity for market participants to circumvent controls in the financial system and create externalities for the rest of society, for example, through facilitating tax evasion or skirting AML laws.

We highlight that there are ways to regulate the DeFi system which would preserve a majority of features of the blockchain architecture but support accountability and regulatory compliance. These solutions would rely on a system where validators on the blockchain agree to check if a particular address belongs to a certified entity and validators would be charged with only processing transactions that involve certified addresses.

How this system evolves in terms of technology and regulation has important consequences for liquidity and credit provision in the economy, and ultimately the standing of the United States and other global economies. There are also strategic and competitive implications across countries. The United States obtains significant economic and strategic benefits from the central role that the US dollar and the US financial system hold internationally. Therefore, it is in the United States' interest to encourage innovation and modern financial technologies but at the same time to set standards that protect consumers and maintain the transparency, accountability, and

stability of the system. The cross-jurisdictional structure of permissionless blockchain ledgers entails a danger that participants will engage in regulatory arbitrage which could undermine the financial system and its stability. Coordination between the main financial markets will be important to prevent a hollowing-out of financial regulations.

ACKNOWLEDGMENTS We thank Janice Eberly, Gary Gorton, Eswar Prasad, Daniel Ferreira, and James Stock for very helpful comments. We also thank Jiageng Liu, Yuou Wu, and Xin Xiong for excellent research assistance.

References

Allen, Franklin, Elena Carletti, and Xian Gu. 2019. "The Roles of Banks in Financial Systems." In *The Oxford Handbook of Banking,* edited by Allen N. Berger, Philip Molyneux, and John O. S. Wilson. Oxford: Oxford University Press.

Aoyagi, Jun. 2020. "Liquidity Provision by Automated Market Makers." Working Paper. Social Science Research Network, October 15. https://ssrn.com/abstract=3674178.

Aoyagi, Jun, and Yuki Ito. 2021. "Coexisting Exchange Platforms: Limit Order Books and Automated Market Makers." Working Paper. Social Science Research Network, March 22. https://ssrn.com/abstract=3808755.

Aramonte, Sirio, Wenqian Huang, and Andreas Schrimpf. 2021. "DeFi Risks and the Decentralisation Illusion." *BIS Quarterly Review* (December): 21–36.

Auer, Raphael, and Stijn Claessens. 2020. "Cryptocurrency Market Reactions to Regulatory News." Working Paper 381. Dallas: Federal Reserve Bank of Dallas, Globalization Institute. https://doi.org/10.24149/gwp381.

Bartlett, Robert, and Eric Talley. 2017. "Law and Corporate Governance." In *The Handbook of the Economics of Corporate Governance, Volume 1*, edited by Benjamin Hermalin and Michael Weisbach. Amsterdam: North Holland.

Beniiche, Abdeljalil. 2020. "A Study of Blockchain Oracles." ArXiv:2004.07140. Ithaca, N.Y.: Cornell University.

Berg, Tobias, Manju Puri, and Jörg Rocholl. 2020. "Loan Officer Incentives, Internal Rating Models, and Default Rates." *Review of Finance* 24, no. 3: 529–78.

Berle, Adolf, and Gardiner Means. 1932. *The Modern Corporation and Private Property*. Piscataway, N.J.: Transaction Publishers.

Biais, Bruno, Christophe Bisière, Matthieu Bouvard, and Catherine Casamatta. 2019. "The Blockchain Folk Theorem." *Review of Financial Studies* 32, no. 5: 1662–715.

Bier, Jonathan. 2021. *The Blocksize War: The Battle for Control over Bitcoin's Protocol Rules.* Independently published.

Breidenbach, Lorenz, Christian Cachin, Benedict Chan, Alex Coventry, Steve Ellis, Ari Juels, and others. 2021. "Chainlink 2.0: Next Steps in the Evolution of Decentralized Oracle Networks." White Paper. https://chain.link/whitepaper.

Budish, Eric. 2018. "The Economic Limits of Bitcoin and the Blockchain." Working Paper 24717. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24717.

Buterin, Vitalik. 2014. "Ethereum: A Next Generation Smart Contract and Decentralized Application Platform." White Paper. https://ethereum.org/en/whitepaper/.

Buterin, Vitalik. 2021. "Moving beyond Coin Voting Governance." Blog post, Vitalik Buterin's Website, August 16. https://vitalik.ca/general/2021/08/16/voting3.html.

Caldarelli, Giulio, and Joshua Ellul. 2021. "The Blockchain Oracle Problem in Decentralized Finance—A Multivocal Approach." *Applied Sciences* 11, no. 16: 7572.

Campbell, John Y. 2016. "Restoring Rational Choice: The Challenge of Consumer Financial Regulation." *American Economic Review* 106, no. 5: 1–30.

Capponi, Agostino, and Ruizhe Jia. 2021. "The Adoption of Blockchain-Based Decentralized Exchanges." ArXiv:2103.08842. Ithaca, N.Y.: Cornell University.

Célérier, Claire, and Boris Vallée. 2017. "Catering to Investors through Security Design: Headline Rate and Complexity." *Quarterly Journal of Economics* 132, no. 3: 1469–508.

Cong, Lin William, and Zhiguo He. 2019. "Blockchain Disruption and Smart Contracts." *Review of Financial Studies* 32, no. 5: 1754–97.

Cong, Lin William, Zhiguo He, and Jiasun Li. 2021. "Decentralized Mining in Centralized Pools." *Review of Financial Studies* 34, no. 3: 1191–235.

Cong, Lin William, and Yizhou Xiao. 2021. "Categories and Functions of Crypto-Tokens." In *The Palgrave Handbook of FinTech and Blockchain,* edited by Maurizio Pompella and Roman Matousek. New York: Springer Nature.

Copeland, Tim. 2020. "Steem vs Tron: The Rebellion against a Cryptocurrency Empire." Decrypt, August 18. https://decrypt.co/38050/steem-steemit-tron-justin-sun-cryptocurrency-war.

De, Nikhilesh. 2021. "Circle Reveals Assets Backing USDC Stablecoin." CoinDesk, July 20. https://www.coindesk.com/business/2021/07/20/circle-reveals-assets-backing-usdc-stablecoin/.

De, Nikhilesh, and Marc Hochstein. 2021. "Tether's First Reserve Breakdown Shows Token 49% Backed by Unspecified Commercial Paper." CoinDesk, May 13. https://www.coindesk.com/markets/2021/05/13/tethers-first-reserve-breakdown-shows-token-49-backed-by-unspecified-commercial-paper/.

Di Salvo, Mat. 2020. "Billions in Ethereum at Play: DeFi Meme Coins Are No Joke." Decrypt, September 4. https://decrypt.co/40939/defi-meme-coins-no-joke-billions-ethereum.

Ferreira, Daniel, Jin Li, and Radoslawa Nikolowa. 2019. "Corporate Capture of Blockchain Governance." Working Paper 593. Brussels: European Corporate Governance Institute (ECGI).

Gorton, Gary B. 2021. "The Orkney Slew and Central Bank Digital Currencies." Working Paper. Social Science Research Network, October 7. https://ssrn.com/abstract=3937323.

Gorton, Gary B., and Jeffery Zhang. 2021. "Taming Wildcat Stablecoins." Working Paper. Social Science Research Network, July 19. https://ssrn.com/abstract=3888752.

Gushue, Joseph P. 2021. "Uniswap v3 Employs a 'New' License Agreement to Stake Copycat Vampire Attacks." Blog post, Imagine That IP Law, Volpe Koenig, April 20. https://www.vklaw.com/ImagineThatIPLawBlog/uniswap-v3-employs-a-new-license-agreement.

Halaburda, Hanna, Zhiguo He, and Jiasun Li. 2021. "An Economic Model of Consensus on Distributed Ledgers." Working Paper 29515. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29515.

Harvey, Campbell R., Ashwin Ramachandran, and Joey Santoro. 2021. *DeFi and the Future of Finance.* Hoboken, N.J.: John Wiley and Sons.

Hermalin, Benjamin, and Michael Weisbach, eds. 2017. *The Handbook of the Economics of Corporate Governance, Volume 1.* Amsterdam: North Holland.

Huberman, Gur, Jacob D. Leshno, and Ciamac Moallemi. 2021. "Monopoly without a Monopolist: An Economic Analysis of the Bitcoin Payment System." *Review of Economic Studies* 88, no. 6: 3011–40.

Kereiakes, Evan, Do Kwon, Marco Di Maggio, and Nicholas Platias. 2019. "Terra Money: Stability and Adoption." White Paper. https://assets.website-files.com/611153e7af981472d8da199c/618b02d13e938ae1f8ad1e45_Terra_White_paper.pdf.

Laibson, David, Andrea Repetto, and Jeremy Tobacman. 2007. "Estimating Discount Functions with Consumption Choices over the Lifecycle." Working Paper 13314. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w13314.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 2000. "Investor Protection and Corporate Governance." *Journal of Financial Economics* 58, no. 1: 3–27.

Leclair, Dylan, and Sam Rule. 2021. "How Leverage in the Derivatives Market Dipped the Bitcoin Price." *Bitcoin Magazine*, November 17. https://bitcoinmagazine.com/markets/leverage-in-derivates-market-dipped-bitcoin-price.

Lehar, Alfred, and Christine A. Parlour. 2021. "Decentralized Exchanges." Working Paper. Social Science Research Network, August 16. https://ssrn.com/abstract=3905316.

Lim, Michelle. 2021. "Iron Finance's DeFi Bank Run—and How Mark Cuban Got 'Rekt.'" Forkast, June 22. https://forkast.news/iron-finances-defi-bank-run-why-mark-cuban-got-rekted/.

Lusardi, Annamaria, and Olivia S. Mitchell. 2007. "Financial Literacy and Retirement Preparedness: Evidence and Implications for Financial Education." *Business Economics* 42, no. 1: 35–44.

Makarov, Igor, and Antoinette Schoar. 2021. "Blockchain Analysis of the Bitcoin Market." Working Paper 29396. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29396.

McFarlane, Gary. 2021. "How Crypto Derivatives Crashed the Market—Do Prices Signal Buy Bitcoin?" Inside Bitcoins, June 3. https://insidebitcoins.com/news/how-crypto-derivatives-crashed-the-market-do-prices-signal-buy-now.

Merton, Robert C. 1995. "A Functional Perspective of Financial Intermediation." *Financial Management* 24, no. 2: 23–41.

Phillips, Todd, and Alexandra Thornton. 2022. "Congress Must Not Provide Statutory Carveouts for Crypto Assets." Center for American Progress, March 1. https://www.americanprogress.org/article/congress-must-not-provide-statutory-carveouts-for-crypto-assets/.

Platt, Moritz, Johannes Sedlmeir, Daniel Platt, Paolo Tasca, Jiahua Xu, Nikhil Vadgama, and Juan Ignacio Ibañez. 2021. "Energy Footprint of Blockchain

Consensus Mechanisms beyond Proof-of-Work." ArXiv:2109.03667. Ithaca, N.Y.: Cornell University.

Potter, Sam. 2022. "Wild Crypto Leverage Is on Offer for Pros in 20-Times Bitcoin Bet." Bloomberg, January 26. https://www.bloomberg.com/news/articles/2022-01-26/wild-crypto-leverage-on-offer-for-pros-in-20-times-bitcoin-bet#xj4y7vzkg.

Prasad, Eswar S. 2021. *The Future of Money: How the Digital Revolution Is Transforming Currencies and Finance.* Cambridge, Mass.: Harvard University Press.

Roe, Mark J. 2018. "Stock Market Short-Termism's Impact." Working Paper. Social Science Research Network, October 22. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3171090.

Ru, Hong, and Antoinette Schoar. 2016. "Do Credit Card Companies Screen for Behavioral Biases?" Working Paper 22360. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w22360.

Schär, Fabian. 2021. "Decentralized Finance: On Blockchain- and Smart Contract-Based Financial Markets." *Federal Reserve Bank of St. Louis Review* 103, no. 2 (April): 153–74.

Stevens, Robert. 2020. "Vitalik Buterin: DeFi Yield Farmers Go Brrr More Than Central Banks." Decrypt, August 31. https://decrypt.co/40318/vitalik-buterin-ethereum-defi-yield-farmers-central-banks.

Szabo, Nick. 1996. "Smart Contracts: Building Blocks for Digital Markets." https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart_contracts_2.html.

Tunggal, Abi Tyas. 2022. "The 65 Biggest Data Breaches (Updated June 2022)." Blog post, UpGuard, June 26. https://www.upguard.com/blog/biggest-data-breaches.

Van Wirdum, Aaron. 2016. "Ethereum Classic Community Navigates a Distinct Path to the Future." *Bitcoin Magazine*, August 19. https://bitcoinmagazine.com/business/ethereum-classic-community-navigates-a-distinct-path-to-the-future-1471620464.

Vigna, Paul. 2022. "Crypto Thieves Get Bolder by the Heist, Stealing Record Amounts." *Wall Street Journal*, April 22. https://www.wsj.com/articles/crypto-thieves-get-bolder-by-the-heist-stealing-record-amounts-11650582598.

Werbach, Kevin. 2018. *The Blockchain and the New Architecture of Trust*. Cambridge, Mass.: MIT Press.

Werbach, Kevin, and Nicolas Cornell. 2017. "Contracts Ex Machina." *Duke Law Journal* 67:313–82.

Wright, Aaron, and Primavera De Filippi. 2015. "Decentralized Blockchain Technology and the Rise of Lex Cryptographia." Working Paper. Social Science Research Network, March 20. https://ssrn.com/abstract=2580664.

Comments and Discussion

COMMENT BY

GARY GORTON[1] Decentralized finance (DeFi) is a blockchain-based set of smart contracts that executes financial transactions without a centralized authority. It relies on member agents jointly making decisions. It is a large and growing sector of crypto space that has the potential to significantly disrupt the financial sector. "Disruptive" in the sense of Christensen (2011), it is a new technology that will reduce or eliminate the need for some set of skills or technologies. For example, the advent of personal computers disrupted the typewriter market. So, the question is: Will DeFi significantly disrupt banking?

Regulators and academics need to understand this space because while DeFi is only embryonic currently, it will grow and morph. Makarov and Schoar clearly and comprehensively summarize the ongoing developments, giving us an introduction to this space. Their overview is important because there is bewildering terminology that is little understood by many.

It is important to keep in mind that we are in the very early days of blockchain, DeFi, smart contracts, and stablecoins. These early days are somewhat like these two examples: in 1899, there were 30 American car companies, and by the end of the next decade an additional 485 had started up. But this number dropped from 253 in 1908 to only 44 by 1929 and three companies—Ford, General Motors, and Chrysler—accounted for

1. Gary Gorton is the Frederick Frank Class of 1954 Professor of Finance at the Yale School of Management. He was a cofounder and board member of TNB USA Inc. (The Narrow Bank), which was a Connecticut-based depository institution. The views expressed in the comment are those of the discussant and do not necessarily reflect those of Yale School of Management.

Brookings Papers on Economic Activity, Spring 2022: 197–215

80 percent of cars produced.[2] Much of the activity in crypto space is akin to this type of tournament competition, where only a handful will survive.

A second example has to do with network effects. The internet began in 1969 when ARPANET was started by the Defense Advanced Research Projects Agency (DARPA), a part of the Department of Defense. By the 1970s there were many networks, but they were not interoperable, that is, they could not talk to each other, so to speak. Technological progress in the form of packet switching, a way to group data so that they can be transmitted over a digital network, allowed the networks to be connected.[3] Currently, there are about 1,000 blockchains, which are not (very) interoperable.[4] And the blockchains are not very scalable. Nor is crypto space interoperable with the current financial sector, including banks and the payment system. But all this will change.

Today, DeFi is more of a promise than a reality. Many DeFi activities vaguely resemble what banks do, like borrowing and lending, trading derivatives, and trading cryptocurrencies (foreign currencies in the case of banks). But these activities take place in a narrowly restricted area: crypto space. Agents lend cryptocurrencies for interest and trade derivatives on Bitcoin, for example. All these activities can also be done on crypto exchanges, where loans and derivative positions must be collateralized and face margin.

Compared to crypto exchanges, it is argued that the promise of DeFi is that there is no central authority. Rather, there is a governance protocol, as described by Makarov and Schoar. In fact, the founding premise of DeFi is that the decentralized nature of DeFi governance avoids agency problems. There is no central authority that must be incentivized and monitored. For DeFi to grow, this avoidance of agency problems must be very efficient. I discuss this below.

Can DeFi disrupt the banking system? Banks intermediate between borrowers and lenders. Banks issue short-term debt and use the proceeds to make loans. Real investment projects are financed by banks. Currently, only one large DeFi platform has its own native currency. MakerDAO has a currency called Dai, which I discuss below. There are currently no loans made off-chain, though this could change in the future.

2. History, "Automobile History," https://www.history.com/topics/inventions/automobiles.

3. See Mowery and Simcoe (2002) and Roberts (1978).

4. See EarthWeb, "How Many Blockchains Are There in 2022?," https://earthweb.com/how-many-blockchains-are-there/#:~:text=Currently%2C%20there%20are%20at%20least,platforms%20provided%20in%20this%20industry.

For a DeFi platform to become a bank, it would have to be interoperable with the existing payments system and have a bank charter giving it access to a Federal Reserve master account.[5] Currently, the only way to link to the world outside cyber space is via an oracle, as explained by Makarov and Schoar. An oracle allows a smart contract to retrieve data from outside crypto space. But to make loans, money must be transferred to firms, requiring stablecoins to become more sophisticated. And to make credit decisions, bit data must be imported and combined with AI and smart contracts. This is already occurring.

In what follows, I review these issues.

DEFI AND AGENCY PROBLEMS Eliminating agency problems in firms is an admirable goal. The root of agency problems is the impossibility of writing a complete state-contingent executive compensation contract. Outcomes of executive actions are noisy. There are measurability problems and asymmetric information. The decisions to be made are not known in advance. Incentives are not aligned between principal and agent. Can technology overcome these problems?

A large body of literature on executive compensation is aimed at understanding whether contracts are useful in mitigating agency problems.[6] The explicit premise of DeFi is that it avoids agency problems since there is no central authority. Instead, there is a decentralized governance mechanism. Key questions are: Can the decentralized governance mechanism work if a DeFi gets large? And does it currently work the way it is designed to work? In other words, can agency problems be eliminated?

Agency problems have been the dominant paradigm in corporate finance since the 1970s, first articulated by Alchian and Demsetz (1972) and Jensen and Meckling (1976). The paradigm goes like this: There is a separation of ownership by shareholders and control by managers. Managers are entrenched and often act in their own self-interest. This is costly. Managers can divert corporate resources to themselves. For example, Dyck and Zingales (2004) examined the premium attached to blocks of stock when they traded. The premium represents the value of the private benefits that will accrue to the owner of the block. "Based on 393 control transactions between 1990 and 2000 we find that on average corporate control is worth 14 percent of the equity value of a firm" (538). This premium is viewed

5. So far, the Federal Reserve has resisted granting fintech banks master accounts; see Hill (2022).

6. See Edmans, Gabaix, and Jenter (2017) for a summary.

as the value of the private benefits that can be extracted by a blockholder; in other words, it is the cost of the agency problem.

As explained by Makarov and Schoar, DeFi is governed directly by shareholders (nodes) via governance tokens. The claim is that there are no agency problems. But while there may be agency costs to an entrenched management, there are also benefits to having a manager. For example, Bennedsen, Pérez-Gonzáles, and Wolfenzon (2020) study the impact of CEOs on performance by analyzing the effect of CEO deaths and the deaths of CEOs' immediate family members: "we find that CEOs' (but not board members') deaths and deaths in CEOs' families are strongly correlated with declines in firm operating profitability, investment, and sales growth" (1877).

Currently, DeFi platforms are not so large and complicated, so their current governance system may well work. But these platforms can grow by becoming larger or by having an increased number of separate DeFi platforms. In either case, it is not clear that decentralized governance will work. There are issues of expertise and attention.

Indeed, it is not clear that decentralized governance currently works the way its proponents claim. Sun, Stasinakis, and Sermpinis (2022) collected information for the MakerDAO ("Maker") protocol performance, including all voters, their choices, and votes in Maker governance polls from August 2019 to October 2021. MakerDAO is one of the oldest and most influential DeFi platforms. They conclude that "by examining Maker governance polls, we find that voters are centralized in a small group, and voting power is unequally distributed among these voters. In most voting activities, the largest voters could account for a significant proportion of votes" (15). This is also consistent with Azouvi, Maller and Meiklejohn (2019), who argue that a few key developers have unilateral decision-making power in blockchain governance.

These findings, albeit only two studies, suggest that agency problems arise endogenously. Some participants become the agents and the others become principals. Further, there are good reasons why we delegate to agents. Agents have more expertise than other participants do, and agents' sole job is to pay attention to the task, whereas other participants have no or little expertise and not enough time to pay close attention.

There are, however, proposals to make different kinds of governance tokens, say different colors, by which some agents are designated "experts" and only they can vote on some decisions (Kaal 2021). Of course, this just pushes the problem down a layer. Who decides who the experts are? Who decides which decisions only the experts can vote on? Having a group of

anonymized individuals who decide to follow a certain protocol to make decisions does not seem like a viable alternative to corporate governance for large, complicated, organizations.

Of course, there are other forms of ownership besides that of the residual claimants, the shareholders. Cooperatives and mutual associations are ownership forms that more closely resemble DeFi organizations (Hansman 1996). When do cooperatives succeed? Holmström (1999): "Collective decision making is always difficult. But it is more difficult the more the interests of the parties diverge. A group with common interests will have a much easier time to reach a good decision than a group with highly divergent interests" (407). The plywood cooperatives in the Northwest of the United States have been around for seventy years (Craig and Pencavel 1992). But in the case of DeFi, many of the major platforms are essentially controlled by the core teams who developed the DeFi. For example, approximately 46 percent of the DeFi Compound's tokens were distributed to shareholders, founders, and the Compound team (Dale 2020). But that leaves 54 percent for the Compound developers.

WILL DEFI DISRUPT THE BANKING SYSTEM? Banks create short-term debt that earns a convenience yield, and they lend out the proceeds to corporate or retail borrowers. These are the core activities of banks. Currently, DeFi activities are self-referential, meaning that all the activities occur in crypto space using a cryptocurrency. Borrowing and lending refers to lending a cryptocurrency to use for other purposes, like buying another cryptocurrency. Lending is at about $35 billion currently, compared to $180 billion in the stablecoin market.[7] DeFi derivatives are used to hedge the price risk associated with a cryptocurrency or to exploit speculative opportunities. These activities have nothing to do with the "real world," and many of these activities are essentially zero-sum games.

Stablecoin issuers, however, are banks. Stablecoins are digital tokens residing on a blockchain that their issuers say are backed one-to-one with cash or safe assets. And their terms of service say that they are redeemable at par on demand. In other words, stablecoin issuers are banks (Gorton and Zhang 2021). And their peg to the dollar holds about 85 percent of the time (Gorton, Ross, and Ross 2022). But stablecoin issuers are not DeFi organizations.

There is one DeFi platform that issues a stablecoin—MakerDAO—and the coin is called Dai.[8] MakerDAO offers a smart contract where a user

7. DeFi Pulse, https://www.defipulse.com/.

8. MakerDAO, "The Maker Protocol: MakerDAO's Multi-Collateral Dai (MCD) System," https://makerdao.com/en/whitepaper.

deposits collateral, which can be any Ethereum-based cryptocurrency, for a loan of an equivalent value of Dai. Maker maintains Dai at one dollar using a system of collateral and price feeds, managed by the MKR token holders; these are the governing agents. Dai has traded at a dollar only 53 percent of the time, whereas other top stablecoins hold their one-to-one peg about 85 percent of the time (Gorton, Ross, and Ross 2022). It is not clear why Dai cannot hold the one-to-one peg.

To be a bank means the entity issues money that has a convenience yield and makes loans to real sector firms. Stablecoins currently are of limited use. You cannot buy your groceries with stablecoins. In fact, stablecoins have a negative convenience yield—an *inconvenience yield* (Gorton, Ross, and Ross 2022). This will change as technology makes blockchains more scalable and interoperable. But stablecoin issuers do make loans, by holding commercial paper. According to Kim (2022), a one standard deviation change in the daily issuance (= $330 million) of the top three stablecoins results in a 7 percent increase in commercial paper issuance the next day (= $198 million). So, although stablecoins might appear to be small compared to the regulated banking system, they are large enough to move money markets.

Lending by DeFi organizations could be made via smart contracts using big data.[9] But the crucial input is the big data. Ant Group, for example, has used big data in many forms in its financial ecosystem. One notable example is its "310 lending model" for making loans to small businesses. The "310" stands for a three-minute application, one-second approval, and zero human intervention. In five years, the loan business grew to $103.4 billion (Ding and others 2018). The algorithm for doing credit analysis is AI based on big data from the use of Taobao, a large online shopping platform. Data are the key here. Di Maggio, Ratnadiwakara, and Carmichael (2022) study one fintech lending platform: "Comparing actual outcomes of the fintech platform's model to counterfactual outcomes based on a 'traditional model' used for regulatory reporting purposes, we find that the latter would result in a 70% higher probability of being rejected and higher interest rates for those approved" (abstract).

DeFi organizations could make loans like banks, but they do not have access to any big data, especially proprietary data. This raises the question of whether DeFi can truly be self-contained, so to speak. Oracles are a way for a DeFi platform to reach out to the world to get data, like current interest rates or FX rates, as discussed by Makarov and Schoar. But today this

9. An example is Aave; see Meegan and Koens (2021).

appears very limited. In principle, data could be imported through an oracle to make lending decisions, but what data? In addition, the DeFi platform would have to be able to analyze the data with machine learning. Currently, there is work being done to combine AI with DeFi.[10]

INTEROPERABILITY For disruption of the financial system, DeFi platforms would have to be interoperable in two senses: (1) blockchain to blockchain, and (2) DeFi platform or stablecoin issuer to the existing payments system. I will focus on the second sense of interoperability.[11]

To be a bank, there must be a link to the existing payments system and to brick-and-mortar banks. In other words, the bank must have a Federal Reserve master account. In May 2021, the Federal Reserve issued a proposal outlining some criteria for who could obtain a master account.[12] As would be expected, there was the usual stuff about financial stability and so on, but there were no real details about any criteria.[13]

Nevertheless, it is hard to stop the tsunami of innovation. Physical cash, paper notes, and coins have been the key medium of exchange in conventional financial systems for a long time. But now it seems clear that blockchain-based systems are here to stay. Firms are adjusting to this new reality. One example is Visa, which recently announced that it would accept the stablecoin USD Coin (Christodorescu and others 2021).[14]

One can imagine banks being dramatically altered. Liabilities will be stablecoins and lending will be through platforms. Clearing and settlement, as it is currently constituted, will be eliminated. International payments will not go through inefficient chains of correspondent banks. Remittances will be much cheaper.

CONCLUSION Blockchain is and will be disruptive. DeFi will probably have limited use. And DeFi is borderless, so it will be difficult to regulate. "Banking" and "money" are changing their forms. But fundamentally their economic structures remain the same. Banks produce money, which is short-term debt redeemable on demand at par. How then can these new technologies be disruptive? We will see.

10. See Badruddoja and others (2021), Sadman and others (2022), and Ouyang, Yuan, and Wang (2015).

11. On the first sense, see Buterin (2016).

12. Federal Reserve System, "Proposed Guidelines for Evaluating Account and Services Requests," https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20210505a1.pdf.

13. For a sample of the controversy, see "Senate Banking Hearing: 'The Semiannual Monetary Policy Report to the Congress.'" https://www.youtube.com/watch?v=cQ2p0FyQVqg.

14. See Visa, "Visa Becomes First Major Payments Network to Settle Transactions in USD Coin (USDC)," March 29, 2021, https://usa.visa.com/about-visa/newsroom/press-releases.releaseId.17821.html.

REFERENCES FOR THE GORTON COMMENT

Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review* 62, no. 5: 777–95.

Azouvi, Sarah, Mary Maller, and Sarah Meiklejohn. 2019. "Egalitarian Society or Benevolent Dictatorship: The State of Cryptocurrency Governance." In *Financial Cryptography and Data Security*, edited by Aviv Zohar, Ittay Eyal, Vanessa Teague, Jeremy Clark, Andrea Bracciali, Federico Pintore, and Massimiliano Sala. Berlin: Springer.

Badruddoja, Syed, Ram Dantu, Yanyan He, and Kritagya Upadhayay. 2021. "Making Smart Contracts Smarter." In *2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. New York: IEEE.

Bennedsen, Morten, Francisco Pérez-Gonzáles, and Daniel Wolfenzon. 2020. "Do CEOs Matter? Evidence from Hospitalization Events." *Journal of Finance* 75, no. 4: 1877–911.

Buterin, Vitalik. 2016. *Chain Interoperability*. R3. https://www.r3.com/wp-content/uploads/2017/06/chain_interoperability_r3.pdf.

Christensen, Clayton M. 2011. *The Innovator's Dilemma: The Revolutionary Book That Will Change the Way You Do Business*. New York: HarperCollins.

Christodorescu, Mihai, Erin English, Wanyun Catherine Gu, David Kreissman, Ranjit Kumaresan, Mohsen Minaei, Srinivasan Raghuraman, Cuy Sheffield, Arjuna Wijeyekoon, and Mahdi Zamani. 2021. "Universal Payment Channels: An Interoperability Platform for Digital Currencies." ArXiv:2109.12194v2. Ithaca, N.Y.: Cornell University.

Craig, Ben, and John Pencavel. 1992. "The Behavior of Worker Cooperatives: The Plywood Companies of the Pacific Northwest." *American Economic Review* 82, no. 5: 1083–105.

Dale, Brady. 2020. "Compound's Approach to DeFi Governance Starts with Giving Away COMP Tokens." CoinDesk, May 28. https://www.coindesk.com/tech/2020/05/28/compounds-approach-to-defi-governance-starts-with-giving-away-comp-tokens/.

Di Maggio, Marco, Dimuthu Ratnadiwakara, and Don Carmichael. 2022. "Invisible Primes: Fintech Lending with Alternative Data." Working Paper 29840. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29840.

Ding, Ding, Guan Chong, David Lee Kuo Chuen, and Tan Lee Cheng. 2018. "From Ant Financial to Alibaba's Rural Taobao Strategy—How Fintech Is Transforming Social Inclusion." In *Handbook of Blockchain, Digital Finance, and Inclusion*, edited by David Lee Kuo Chuen and Robert Deng. London: Academic Press.

Dyck, Alexander, and Luigi Zingales. 2004. "Private Benefits of Control: An International Comparison." *Journal of Finance* 59, no. 2: 537–600.

Edmans, Alex, Xavier Gabaix, and Dirk Jenter. 2017. "Executive Compensation: A Survey of Theory and Evidence." In *The Handbook of the Economics of*

Corporate Governance, Volume 1, edited by Benjamin E. Hermalin and Michael S. Weisbach. Amsterdam: North-Holland.

Gorton, Gary B., Chase P. Ross, and Sharon Y. Ross. 2022. "Making Money." Working Paper 29710. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29710.

Gorton, Gary B., and Jeffery Zhang. 2021. "Taming Wildcat Stablecoins." Working Paper. Social Science Research Network, July 19. https://ssrn.com/abstract=3888752.

Hansman, Henry. 1996. *The Ownership of Enterprise*. Cambridge, Mass.: Harvard University Press.

Hill, Julie Andersen. 2022. "Bank Access to Federal Reserve Accounts and Payment Systems." Working Paper. Social Science Research Network, March 4. https://ssrn.com/abstract=4048081.

Holmström, Bengt. 1999. "The Future of Cooperatives: A Corporate Perspective." *Finnish Journal of Business Economics* 4:404–17.

Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure." *Journal of Financial Economics* 3, no. 4: 305–60.

Kaal, Wulf A. 2021. "Blockchain-Based Corporate Governance." *Stanford Journal of Blockchain Law and Policy* 4, no. 1. https://stanford-jblp.pubpub.org/pub/blockchain-corporate-governance/release/1.

Kim, Sang Rae. 2022. "How the Cryptocurrency Market Is Connected to the Financial Market." Working Paper. Social Science Research Network, May 13. https://ssrn.com/abstract=4106815.

Meegan, Xavier, and Tommy Koens. 2021. "Lessons Learned from Decentralised Finance (DeFi)." White Paper. Amsterdam: ING Bank. https://www.ingwb.com/binaries/content/assets/insights/themes/distributed-ledger-technology/defi_white_paper_v2.0.pdf.

Mowery, David C., and Timothy Simcoe. 2002. "Is the Internet a U.S. Invention?—An Economic and Technological History of Computer Networking." *Research Policy* 31, nos. 8–9: 1369–87.

Ouyang, Liwei, Yong Yuan, and Fei-Yue Wang. 2015. "Learning Markets: An AI Collaboration Framework Based on Blockchain and Smart Contracts." *Internet of Things Journal* 14, no. 8: 1–14.

Roberts, Lawrence. 1978. "The Evolution of Packet Switching." *Proceedings of the IEEE* 66, no. 11: 1307–13.

Sadman, Nafiz, Md Manjurul Ahsan, Abdur Rahman, Zahed Siddique, and Kishor Datta Gupta. 2022. "Promise of AI in DeFi, a Systematic Review." *Digital* 2, no. 1: 88–103.

Sun, Xiaotong, Charalampos Stasinakis, and Georigios Sermpinis. 2022. "Decentralization Illusion in DeFi: Evidence from MakerDAO." ArXiv:2203.16612v1. Ithaca, N.Y.: Cornell University.

COMMENT BY

ESWAR PRASAD In their paper, Makarov and Schoar do a very nice job of discussing the exciting new world of cryptocurrencies and decentralized finance in a rigorous but balanced fashion. Here, I attempt to complement their discussion by discussing some aspects of decentralized finance and the regulation of the new financial ecosystem spawned by these new technologies.

COIN OFFERINGS The proliferation of cryptocurrencies has resulted in the creation of new financial instruments. An initial coin offering (ICO) is a fundraising tool that involves the generation and sale of a set of blockchain-based tokens to finance a particular project or initiative that is usually also blockchain-based. The tokens are sold in exchange for one of the prominent cryptocurrencies or for fiat currencies, and they then become linked to the project they helped finance. An important difference relative to an IPO (an initial public offering of stock for a company listed on a stock exchange) is that an ICO usually does not involve the transfer of ownership stakes to investors.

ICOs are, in effect, bets on the future of a particular cryptocurrency. In the United States, ICOs are far easier to implement than IPOs, which require filings with the Securities and Exchange Commission (SEC) and other extensive disclosure requirements. Companies undertaking ICOs simply create white papers explaining the project's business model, the amount of money they plan to raise (usually a maximum amount is specified), the duration of the ICO campaign, and who is eligible to participate in the ICO. Most ICOs have been carried out on the Ethereum platform.

ICOs have become a key source of funding for blockchain start-ups and other firms operating at the frontiers of this technology. ICOs hold the promise of extraordinary returns for believers in the transformative potential of this technology, but they also imply huge risks for investors. Investors usually have little information beyond a white paper describing the ICO with which to evaluate the business model and the earnings potential of the issuer.

Some ICOs take the form of equity token offerings (ETOs). A company conducting an ETO adds shares to its capital. These shares, which are recorded on a blockchain, grant investors a percentage of voting rights as well as titles of ownership within the company. This differentiates ETOs from normal ICOs, which do not involve any transfer of ownership stakes.

Initial exchange offerings (IEOs) are similar to ICOs except that the tokens are issued through a partnering exchange rather than directly to

investors. The exchange does not in any way guarantee the value or legitimacy of the token issued through an IEO. Still, IEOs are seen as safer than ICOs: the exchange has an incentive to carry out due diligence on the issuing company and its business model, since the exchange faces risk to its reputation if the tokens prove worthless or fraudulent.

IEOs conducted by a particular exchange tend to be standardized, unlike ICOs, whose terms and structure are determined at the sole discretion of the issuing company. Another difference is that tokens issued through an IEO are immediately tradable on the issuing exchange; with ICOs that is not necessarily the case, especially where there are private placements.

Recognizing that tokenization could be used to broaden the investor base for their offerings, some governments and financial institutions have used another investment product, security token offerings (STOs), which in some ways bridges the gap between IPOs and their cryptocurrency counterparts. STOs involve selling digital tokens on cryptocurrency exchanges. Security tokens are securities, similar to stocks and bonds, that usually represent ownership stakes in a particular company. The tokens represent ownership information about the investment product, recorded on the blockchain. STO tokens are sometimes backed by and represent ownership shares in particular tangible assets, especially illiquid assets such as real estate and fine art. STOs are generally regulated as securities, offering more protection to investors. In the United States, for instance, the SEC has jurisdiction over STOs.

All of these types of digital coin offerings show how blockchain technology is powered by and, in turn is changing finance. Innovations in digital and financial technologies are feeding off each other, creating more opportunities for direct financing of innovative technologies and giving even retail investors the opportunity to participate in the financial benefits (and risks) that could flow from such innovations.

DECENTRALIZED FINANCE Decentralized finance (DeFi) or open finance is a model for providing a broad range of financial services—including credit, savings, and insurance—in a decentralized manner and making the services and products available to anyone in the world (Harvey, Ramachandran, and Santoro 2021; Prasad 2021).

DeFi is built on decentralized blockchains. There are three elements that characterize such systems. Decentralized blockchains have decentralized architectures (no centralized point of failure), decentralized governance (control rests with the members of a network rather than a central authority), and decentralized trust (trust is achieved through a public

consensus mechanism). But the system is logically centralized—the entire network of nodes that make up such a system is linked and is in a commonly agreed-to state at all times. Bitcoin could be considered the earliest form of DeFi.

DeFi relies on smart contract blockchains, of which Ethereum is by far the most widely used. In principle, decentralization confers many advantages over traditional financial systems. One is fault tolerance—failure is less likely because such a system relies on many separate components. Another is attack resistance—there is no central point, such as a major financial institution or centralized exchange, that is vulnerable to attack. A third advantage is collusion resistance—it is difficult for participants in a large decentralized system to collude; corporations and governments, by contrast, have the power to act in ways that might not necessarily benefit common people. A decentralized system is also permissionless (anyone can use it), censorship resistant (no one can stop it), and open (anyone can verify the execution of a transaction).

DeFi has spawned new and creative financial products. For instance, a flash loan is a type of smart contract that typically involves borrowing without collateral, using that money for a transaction, and then returning the borrowed amount, all for a fee that is usually very small. A flash loan is initiated, executed, and completed essentially instantaneously. The key element of a flash loan is that all elements of the contract are executed serially in a batch operation on Ethereum. This eliminates default risk—if the loan is not repaid, the entire set of transactions is nullified. Since it is instantaneous, a flash loan also involves no liquidity risk—if any of the parties in a transaction could not meet their commitments, the flash loan would simply disintegrate, rolling back all of the operations.

A flash loan can be used to arbitrage among assets or across markets without having the principal needed to execute the arbitrage. Such arbitrage behavior can actually make markets more efficient by eliminating price differentials, so flash loans might serve a useful purpose. Flash loans can also be used to refinance loans and other operations that involve swapping various kinds of assets and liabilities.

One of the broader attractions of DeFi is a feature referred to as permissionless composability. This means that a developer can easily, and without having to seek permissions, connect multiple DeFi applications built on open-source technology to create new financial products and services. For example, a user can deposit cryptocurrency into a loan contract, withdraw some stablecoins collateralized by that deposit, and put those stablecoins in a yield-bearing contract. Multiple users pooling their stablecoins could

even build a savings game on top of that structure—all of the interest earned on the pooled stablecoins is awarded to a lucky winner, with everyone else getting their initial deposits back. In principle, compliance tools can also be plugged into such a structure to ensure regulatory compliance in each relevant jurisdiction.

DeFi certainly has the potential to expand the frontier of finance and democratize it. However, while DeFi protocols are already dealing in large amounts of money, there are many questions about whether DeFi operations can be scaled up to rival traditional financial institutions in any serious way.

DeFi diminishes some risks while creating new ones. Since flash loans are instantaneous, default and liquidity risks are minimized. Moreover, computer tools can perform rigorous economic risk assessments of smart contracts and specific DeFi products. Despite the open-source nature of DeFi applications, which should help expose and eliminate security and other weaknesses, there are many residual risks. Sophisticated hackers have been able to take advantage of technical and design vulnerabilities in DeFi products. Malevolent agents can exploit the larger "attack surface" created when combining multiple applications. Other risks that could undermine confidence include software bugs and users who do not fully understand the risks of such products.

Blockchains are self-contained but need information about prices and ownership of assets to execute certain transactions. Computer programs called oracles obtain such off-chain information and also pass on-chain information back to the real world. Oracles are vulnerable to technical risks, including hacks, and to problems with external data providers.

Certain hacks are difficult to thwart because decentralization implies the absence of a central authority to police such behavior or put in place safeguards. DeFi relies on idealistic libertarian norms, such as its own rule of law, with the community creating and enforcing rules that are in the broad interests of stakeholders. In reality, nascent blockchain systems are vulnerable to governance capture by small groups of stakeholders who could twist rules in their favor.

FINANCIAL REGULATION The approaches of governments and central banks to permitting and regulating cryptocurrencies span a wide spectrum. One question for financial regulators is whether there are implications for institutions that fall within their regulatory ambit or if there are any other systemic implications that merit their intervention. Another set of concerns arises regarding whether cryptocurrencies can be used for money laundering, tax evasion, and illicit commerce.

The regulatory responses can be classified into three broad categories. First, a number of countries do not limit the trading or use of cryptocurrencies but are endeavoring to create a framework in which to regulate them and related financial products. The United States regards Bitcoin and other cryptocurrencies as financial assets that are subject to tax laws as well as anti–money laundering (AML) regulations and regulations designed to combat the financing of terrorism (CFT). Canada and Japan have explicit laws concerning the trading and use of cryptocurrencies.

Second, a number of countries have either limited or banned the use of cryptocurrencies altogether. China banned domestic Bitcoin exchanges when it was trying to restrict speculative capital outflows in 2017 and subsequently blocked access to cryptocurrency exchanges. China also banned domestic ICOs, along with prohibiting individuals and institutions from participating in them. In April 2018, India's central bank, the Reserve Bank of India, prohibited banks, financial institutions, and other regulated entities from dealing in virtual currencies, although this was overturned by the country's supreme court in 2020.

A third approach, adopted by the majority of countries, is passive tolerance. This involves not banning cryptocurrencies but discouraging their use by financial institutions and, in many cases, not clarifying the legal status of such currencies even as means of payment. The lack of regulatory clarity often serves as an effective deterrent to the wider use of cryptocurrencies. It stifles innovation as entrepreneurs fear running afoul of the law and discourages investors who lack protection and fear being taken advantage of by unscrupulous operators. Indeed, government oversight can be a powerful tonic in building confidence that cryptocurrencies and related financial products will at least not easily become scams.

The US experience is a useful illustration of the range of financial activities facilitated by cryptocurrencies and the potential for gaps in regulatory oversight to remain as regulators sort through jurisdictional issues.

US law does not yet provide for direct, comprehensive federal oversight of Bitcoin and other cryptocurrencies or the exchanges on which they are traded. State banking regulators oversee certain US and foreign virtual currency spot exchanges largely through state money transfer laws. The Internal Revenue Service treats virtual currencies as property, which means that cryptocurrency holdings have to be reported on income tax filings and they are subject to capital gains taxes. The Treasury's Financial Crimes Enforcement Network (FinCEN) monitors Bitcoin and other virtual currency transfers, focusing on AML/CFT and know your customer requirements.

The SEC has ruled that Bitcoin and Ether are not securities and therefore do not fall under its regulatory purview. If these cryptocurrencies were to be bundled into investment vehicles such as exchange-traded funds, however, they would become traded securities subject to SEC regulation. The SEC also has the authority to oversee ICOs because they typically involve the offer and sale of securities.

The Commodity Futures Trading Commission (CFTC) has declared virtual currencies to be commodities subject to oversight under its authority under the Commodity Exchange Act. Cryptocurrency futures and options fall within its regulatory ambit, but the agency has only limited jurisdiction over spot markets for cryptocurrency trading; it is entitled to act only against fraud, market manipulation, and failure to deliver the commodity.

As Bitcoin and other cryptocurrencies, along with the technologies underpinning them, start playing a bigger role in financial markets, issues of regulatory jurisdiction and the potential for regulatory gaps take on greater significance. One example that illustrates this problem is that the CFTC seems to regulate spot markets for cryptocurrencies and cryptocurrency-related assets mainly through aggressive enforcement. It appears that the agency does not have the power to proactively set standards for spot markets or require dealers to comply with the CFTC's requirements. This is a consistent theme as many of the efforts of regulatory agencies seem to involve interpreting existing statutes and legislation to bring cryptocurrency-related activities into their regulatory ambit rather than developing new standards and statutes that address some of the novel aspects of cryptocurrencies and the financial products they are spawning.

The president's executive order on digital assets, which was issued in March 2022, sets out an ambitious agenda for regulating cryptocurrencies, stablecoins, and blockchain-based finance, potentially giving the United States a key role in defining global standards in these areas. By design, this is a document that provides a comprehensive overview of a path to regulating new financial technologies and products in a manner that allows potential benefits to be realized while mitigating risks to consumers, businesses, and overall financial stability. This still leaves open the difficult challenges of assigning responsibility across agencies for regulating particular products and technologies while also developing specific regulatory policies that balance the needs of facilitating innovation while reducing risks.

Cryptocurrencies may also require greater coordination and harmonization of regulatory efforts across national regulators. While some cryptocurrency exchanges are nominally domiciled in specific countries, the nature

of these virtual currencies makes it difficult to subject them to national rules and regulations, especially with respect to investor protection.

REFERENCES FOR THE PRASAD COMMENT

Harvey, Campbell R., Ashwin Ramachandran, and Joey Santoro. 2021. *DeFi and the Future of Finance*. Hoboken, N.J.: John Wiley and Sons.

Prasad, Eswar S. 2021. *The Future of Money: How the Digital Revolution Is Transforming Currencies and Finance*. Cambridge, Mass.: Harvard University Press.

GENERAL DISCUSSION Robert Hall said that the complexity of decentralized finance (DeFi) will make it difficult to implement clear regulations, and that DeFi practitioners are likely to find workarounds to any regulations that are imposed. He also discussed the similarities between DeFi and existing technologies. For example, he noted that the concept of smart contracts already exists, since lawyers can create legally binding agreements via word processing software. Hall also pointed out that stablecoins are almost identical in their function to money market mutual funds. He explained that runs on money market funds occurred during the global financial crisis because they ignored provisions of the Investment Company Act of 1940, and that stablecoins do not provide any additional benefits compared to well-regulated money market funds. Hall described DeFi as a dead end.

Antoinette Schoar replied that crypto technologies are still in their early development. She mentioned that smart contract platforms have the potential to facilitate new types of transactions and offer increased openness, scale, and simplicity compared to current payment systems. However, she noted that many types of transactions do not need the permissionless and anonymous features of the Bitcoin blockchain, and that the benefits of these technologies could be obtained via a regulated system that addresses their externalities. With regard to smart contracts, Schoar explained that their self-executing nature requires them to be complete contracts ex ante, since there is no method for obtaining ex post remediation via the legal system. She remarked that this offers potential benefits—including as a self-commitment mechanism or to reduce legal costs—but does not allow disadvantaged parties to lodge legal complaints or be made whole if they were defrauded.

Donald Kohn agreed with Hall that the regulation of stablecoins could be dealt with similar to money market funds. Kohn wondered about other potential financial stability issues related to DeFi applications. He asked

the authors whether a decentralized system would have issues with leverage or maturity transformation, which could lead to a run on or fire sale of assets during a flight-to-liquidity episode like the one that occurred in March 2020. Kohn questioned how central banks might intervene in such a scenario.

Igor Makarov responded to the discussants by expanding upon the different variants of stablecoins. He remarked that they mostly focused on stablecoins backed by traditional (and liquid) assets but that so-called algorithmic stablecoins have no associated collateral, posing a potential stability risk. He described past episodes in which large declines in the price of Ethereum led to amplification effects. In these episodes, stablecoins that used Ethereum as collateral—such as the cryptocurrency Dai—saw contracts get executed and unwound, leading to further declines in Ethereum's price. Makarov said that scenarios like this could require central bank interventions to inject liquidity.

Eswar Prasad agreed that these were serious concerns. He noted that because many stablecoins are collateralized—either by liquid assets or other cryptocurrencies—the risks differ from those faced in a traditional fractional reserve banking system. Prasad remarked that financial stability concerns could instead arise via a crisis of confidence. For example, cyberattacks on blockchains could undermine public trust in these systems and incite widespread redemption requests, leading to similar problems to those faced by money market funds during the global financial crisis. As during that episode, Prasad explained, the central bank does not have obvious conduits to funnel liquidity into the system given its decentralized and nontraditional structure.

Janice Eberly asked the panelists about their views on the policy agenda for DeFi technologies and their recommendations for policymakers.

Gary Gorton expressed his agreement with the conclusions of the 2021 President's Working Group on Financial Markets report on stablecoins.[1] He said that stablecoins have resurrected an issue that has long since been decided about whether governments should be the monopoly supplier of money. Gorton commented that given stablecoins' current prevalence and lobbying power, it is too late to ban them. He argued that the next best option is to declare stablecoins as banks that issue short-term debt and to insure them for the maximum amount possible.

1. President's Working Group on Financial Markets, the Federal Deposit Insurance Corporation, and the Office of the Comptroller of the Currency, *Report on Stablecoins*, November 2021, https://home.treasury.gov/system/files/136/StableCoinReport_Nov1_508.pdf.

Prasad expanded upon the potential benefits of cryptocurrency and DeFi. He noted that 5.4 percent of US households are unbanked and that digital payments can help to improve financial inclusion.[2] He added that DeFi technologies offer the potential to improve the current digital payments system in the United States, which he described as inefficient and lagging behind foreign countries like China. In particular, Prasad remarked that DeFi has provided competition to fiat currencies in both its function as a medium of exchange and as a store of value. He concluded that the digital revolution in payments has been successful in forcing policymakers to rethink the current system but that regulators will need to implement appropriate guardrails to prevent financial risks.

Olivier Blanchard asked Gorton about the feasibility of treating stablecoins as banks. He wondered whether the constraints were technical or related to the lobbying power of DeFi companies.

Gorton responded that there was a large lobby and that the current sentiment in Congress is to not limit innovation. He added that stablecoins are difficult to redeem for cash but that sell-off episodes could drop prices to zero. Gorton concluded that this scenario would not have sufficient impact to provoke a financial crisis, but he estimated that it could reach that point within ten years.

Daniel Tarullo asked Gorton whether any other aspects of the DeFi architecture posed risks, and he asked Schoar if the types of governance problems she mentioned could be addressed by the US Securities and Exchange Commission (SEC) rather than by prudential regulators.

Gorton answered that he didn't believe that the leverage associated with borrowing and lending in DeFi apps posed systemic risks. He hypothesized that even if activity in that space ceased entirely, it wouldn't threaten the broader financial system. However, Gorton explained that the short-term debt associated with stablecoins poses a more serious issue. He stated that although their funding is currently only $180–200 billion, activity in stablecoins has influenced the commercial paper market.[3] Gorton expressed his concern that US regulators missed the boat on outlawing stablecoins and that regulations are unlikely to emerge until a future financial crisis.

2. FDIC, "How America Banks: Household Use of Banking and Financial Services," December 17, 2021, https://www.fdic.gov/analysis/household-survey/index.html.

3. Sang Rae Kim, "How the Cryptocurrency Market Is Connected to the Financial Market," working paper, Social Science Research Network, May 13, 2022, https://ssrn.com/abstract=4106815.

Schoar affirmed that certain governance issues could be addressed by the SEC. However, she explained that she views the broader problem as the fact that decentralized autonomous organizations are attempting to operate outside the traditional legal and regulatory systems. Under the current DeFi architecture, she explained, minority shareholders who are defrauded do not benefit from minority shareholder protection or fiduciary duty standards. Schoar also mentioned that the pseudonymous and permissionless aspects of the blockchain could contribute to tax evasion.

PANEL on FISCAL POLICY AND BUDGET DEFICITS FOLLOWING THE PANDEMIC

Editors' Note: This panel, presented at the Spring 2022 *BPEA* conference, discussed the rise in sovereign budget deficits in the United States and around the world and their potential implications for policymaking. N. Gregory Mankiw, of Harvard University, Phillip L. Swagel, of the Congressional Budget Office, and Carmen M. Reinhart, formerly of World Bank Group, participated as panelists. Mankiw and Swagel contributed short papers to this issue based on their remarks. The recording of the panel discussion can be found at https://www.brookings.edu/events/bpea-spring-2022-conference/.

N. GREGORY MANKIW
Harvard University

Government Debt and Capital Accumulation in an Era of Low Interest Rates

ABSTRACT This essay discusses the reasons for and implications of the decline in real interest rates around the world over the past several decades. It suggests that the decline in interest rates is largely explicable from trends in saving, growth, and markups. In this environment, greater government debt is likely not problematic from a budgetary standpoint. But a Ponzi-like scheme of perpetual debt rollover might fail, and such a failure would make a bad state of the world even worse. In addition, even if a perpetual debt rollover succeeds, the increased debt could still crowd out capital, reducing labor productivity, real wages, and consumption.

Everyone has heard the apocryphal Chinese curse, "May you live in interesting times." For better or (mostly) worse, we are living in interesting times. One especially interesting feature of the current macroeconomic environment is the low level of long-term real interest rates. The average historical real return on bonds over the past century is around 250 to 300 basis points, and that is about where real yields stood in the mid-1990s. As I write this essay in March 2022, the yields on US inflation-adjusted bonds of all maturities—even as long as thirty years—are less than zero.

This decline in real interest rates is not unique to the United States but is a worldwide phenomenon. In November 2021, the United Kingdom sold a fifty-year inflation-adjusted bond with a yield of –2.4 percent. That means that bond holders will receive, a half century later, only 30 percent as much purchasing power as they used to buy the security.

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

Brookings Papers on Economic Activity, Spring 2022: 219–231

Like many economists, I have been pondering the causes of the decline in real interest rates and its implications for fiscal policy. I don't pretend to have all the answers. But this brief essay offers a progress report on my thinking.

I. Insights from Neoclassical Growth Theory

The place I would like to begin is with neoclassical growth theory. Of course, monetary policy has a dominant influence on interest rates in the short run. But textbook macroeconomics teaches that monetary policy is neutral in the long run. The downward decline in real interest rates has unfolded over several decades, and the current term structure for inflation-indexed bonds suggests that low real rates will likely persist for at least a few decades more. That sounds like the long run to me. To understand the trend in real interest rates, therefore, my thoughts turn to models of long-run growth, which emphasize investment demand and saving propensities rather than monetary policy.

In particular, by "neoclassical growth theory," I mean the Solow (1956) growth model and the Diamond (1965) overlapping generations model. These models assume certainty and competitive markets, and shortly I will suggest that these assumptions are problematic. But these models are a good starting point, and they offer some useful insights. I will assume that the reader is familiar with them. If you are not, get yourself a copy of David Romer's wonderful textbook, pronto (Romer 2019).

Using conventional notation and assuming a Cobb-Douglas production function, the steady-state real interest rate in the Solow model is given by the equation

$$r = \alpha\left(\frac{n + g + \delta}{s}\right) - \delta,$$

where α is capital's exponent in the production function, n is the rate of population growth, g is the rate of labor-augmenting technological progress, δ is the depreciation rate, and s is the gross saving rate. This equation follows from the model's steady-state condition and the equality of the real interest rate with the net marginal product of capital.

One nice thing about this equation is that it allows us to glean how various changes in the economic environment affect the equilibrium real interest rate. For example, some economists have suggested that the saving rate has increased because rising inequality has shifted income toward

households with higher propensities to save.[1] Others have suggested that the world is experiencing a "saving glut" due to the rapid growth of high-saving economies, such as China.[2] Whatever the reason, other things being equal, a higher saving rate depresses the real interest rate.

How big is this effect? Differentiating the above equation yields

$$\frac{\partial r}{\partial s} = -\alpha\left(\frac{n+g+\delta}{s^2}\right).$$

A plausible calibration is $\alpha = 1/3$, $n = 0.01$, $g = 0.02$, $\delta = 0.05$, and $s = 0.24$, which tells us

$$\frac{\partial r}{\partial s} \approx -(1/3)\left(\frac{.01+.02+.05}{.24^2}\right) = -0.46.$$

Each additional percentage point in the saving rate reduces the steady-state real interest rate by 46 basis points.

The World Bank reports data on the world gross saving rate (as a percentage of gross national income) from 1975 to 2020. It shows a clear upward trend, as seen in figure 1. The world saving rate averaged 25.1 percent during the latter half of this period, compared with 22.2 percent during the first half. An increase in the saving rate of 2.9 percentage points can explain a decline in the real interest rate of about 133 basis points.

Another development, however, is more important. The rate of growth, represented in the Solow model by $n + g$, has declined in recent years, in part due to lower population growth and in part due to lower productivity growth. Again, the Solow model yields a precise answer about how much this change affects the steady-state interest rate:

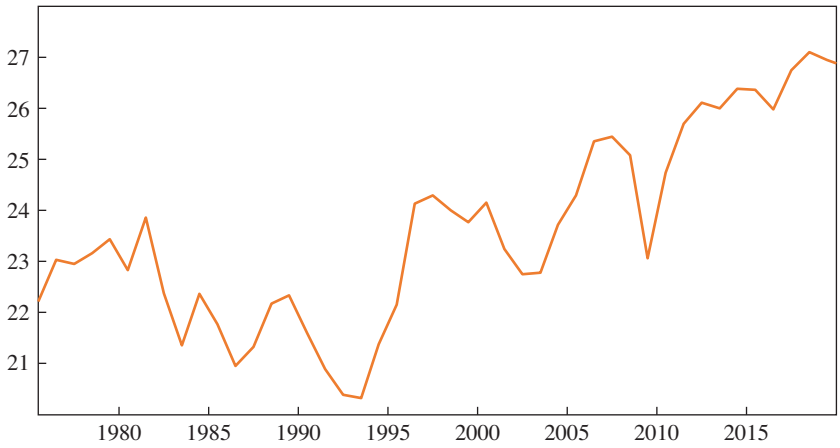
$$\frac{\partial r}{\partial(n+g)} = \frac{\alpha}{s}.$$

With my calibrated parameters, this becomes

$$\frac{\partial r}{\partial(n+g)} \approx \frac{1/3}{0.24} = 1.39.$$

1. See, for example, Straub (2019).
2. See, for example, Bernanke (2005).

Figure 1. World Saving Rate



Source: World Bank.
Note: Data are available at https://data.worldbank.org/indicator/NY.GNS.ICTR.GN.ZS.

A decline in the growth rate of 1 percentage point reduces the real interest rate by 139 basis points.

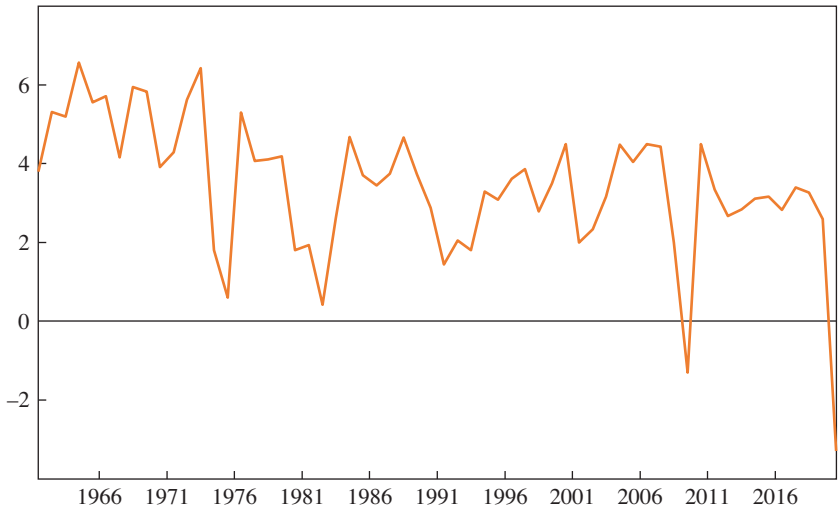
This effect goes a long way toward explaining the decline in interest rates. The World Bank reports data on world GDP growth from 1961 to 2020, shown in figure 2. World GDP growth averaged 2.8 percent per year in the most recent three decades, compared with 4.1 percent per year in the previous three—a fall of 1.3 percentage points. A change of this magnitude can explain a decline in real interest rates of about 181 basis points.

These calculations lead me to conclude that the decline in the real interest rate over the past few decades is not all that mysterious.[3] Based on just the textbook Solow model, the observed higher saving and lower growth rates can together explain a decline in the real interest rate of more than 3 percentage points, which is in the ballpark of what has occurred.

To be sure, this application of the Solow model might strike some readers as audacious or perhaps even foolhardy. The world does not constitute a single economy with fully integrated capital markets. Even if it did, convergence to the Solow steady state may be slow enough that applying steady-state conditions is not fully appropriate. I present these rough calculations not to reach a definitive conclusion but instead to establish proof of concept. Increasing saving and declining growth are powerful forces that have been exerting strong downward pressure on real interest rates

3. Rachel and Smith (2017) reach a similar conclusion.

Figure 2. World GDP Growth



Source: World Bank.
Note: Data are available at https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG.

around the world. Neoclassical growth theory suggests that as long as saving remains high and growth remains low, real interest rates are unlikely to return to historical norms.

What does a low interest rate mean for fiscal policy? To answer this question, we must turn from the Solow growth model to its close cousin, the Diamond overlapping generations model. The Diamond model follows the Solow model in assuming certainty, competitive markets, and a production technology with constant returns to scale in capital and labor and exogenous technological progress. But the Diamond model replaces the assumption of an exogenous saving rate with finitely lived agents who optimize subject to explicit budget constraints. This change permits the incorporation of government debt, so we can examine how debt affects capital accumulation and welfare.

The bottom line from the Diamond model is that the comparison of the real interest rate and the growth rate is crucial (as indeed it is in the Solow model). If the interest rate is less than the growth rate, as seems to be the case today, the economy is in a dynamically inefficient equilibrium. That is, it is saving so much that the capital stock exceeds the level that maximizes steady-state consumption.[4] In this case, the government can run a sustainable

4. Phelps (1961) dubbed this the golden rule level of capital.

Ponzi scheme by issuing debt and rolling it over, along with the accumulating interest, forever.[5] To be sure, the government debt will absorb saving and crowd out capital, but that is a good thing because the economy has too much of it. Government debt can raise welfare when the debt is issued, in the steady state, and along the entire transition path.

At this point, one might think that neoclassical growth theory offers a sanguine view about our current situation of high and growing government debt. And indeed that seems to be the stance of some who have been warning about secular stagnation.[6] They have, at least implicitly, been seeing the world through the lens of neoclassical growth theory. From this perspective, low interest rates are a sign of too much saving and too much capital.[7] This problem can be solved by what might otherwise be considered profligate fiscal policy.

But not so fast. In my discussion so far, I have been sweeping under the rug a notable problem with applying conventional neoclassical growth theory. While it can plausibly explain the decline in real interest rates, it cannot as easily explain the level. Let's return to the equation I started with and plug in my parameters:

$$r = \alpha\left(\frac{n+g+\delta}{s}\right) - \delta \approx \left(\frac{1}{3}\right)\left(\frac{.01+.02+.05}{.24}\right) - .05 = .061.$$

The calibrated Solow model gives a real interest rate of about 6 percent. No plausible set of parameters gives an interest rate less than zero, as we are now observing for long-term inflation-adjusted bonds.

Something must be missing from the model. In fact, two things are missing: risk and market power. Here I consider them separately, though the real world includes risk and market power simultaneously. More research is needed on their possible interactions.

II. Adding Risk

Let's first consider risk. In the Diamond model, the real interest rate on government bonds equals the net marginal product of capital. But that is not true in the world. Government bonds are safe, whereas growth and capital ownership are risky. A risk premium separates the return on safe assets from the return on capital.

5. For more on this topic, see O'Connell and Zeldes (1988).
6. See, for example, Summers (2016).
7. This situation of low real interest rates also makes the zero lower bound a more frequent constraint on monetary policy, but I won't address that topic here.

An increase in the risk premium can drive down the safe interest rate, and it is possible that a rising risk premium can help explain the observed decline in real interest rates. Gauging this effect is difficult because changes in risk premiums are hard to measure. But I doubt that a rising risk premium is an important part of the story. Stock market valuations, such as price–earnings ratios, have risen while real interest rates have fallen, suggesting that the expected return on risky assets has fallen as well. My best guess is that a rising risk premium does not explain the decline in real interest rates, though the existence of a risk premium is one reason real rates are always low compared with the return on capital.

A small body of literature reexamines the issues of dynamic efficiency, capital accumulation, and government debt in environments with uncertainty. Many years ago, I wrote a paper on this topic with Andy Abel, Larry Summers, and Richard Zeckhauser (1989) and then another one with Larry Ball and Doug Elmendorf (1998). Olivier Blanchard's American Economic Association Presidential Address in 2019 has renewed interest in the subject.[8] This literature has not settled all the issues, but let me summarize what I believe to be true.

First, comparing an economy's safe interest rate with its average growth rate does not reveal anything about its dynamic efficiency. Uncertainty generates a risk premium, which depresses the safe interest rate. Economies that are efficient in every way can have low safe rates of interest if risk and risk aversion are high enough.

Second, judging the efficiency of capital accumulation is harder in economies with uncertainty, but it is not impossible. Abel, Summers, Zeckhauser, and I (1989) proposed a criterion for overlapping generations models with uncertainty: if the cash flow earned by capital always exceeds the cash flow used for capital investment, the economy is efficient.[9] That criterion appears to be satisfied in actual economies.[10]

8. Peterson Institute for International Economics, https://www.piie.com/commentary/speeches-papers/public-debt-and-low-interest-rates.

9. The efficiency criterion in Abel and others (1989) establishes a form of Pareto optimality: no person can be made better off without someone else being made worse off. But note that we define a person to be someone born in a particular time and a particular state of nature. This approach precludes some welfare improvements from intergenerational risk sharing. These could require a person to be born at a particular time to evaluate her situation as of time zero, recognizing the various states of nature that might occur when she is born. In a sense, a person in pre-birth limbo must be willing to trade off welfare among different possible versions of herself.

10. Because the condition presented in Abel and others (1989) appears to be satisfied in the real world, my subsequent work on this topic typically restricts itself to theoretical frameworks in which this condition holds. That is not true of all work in this literature. For example, this condition does not hold in the example emphasized in Blanchard (2019).

Third, if the government in a dynamically efficient economy observes a safe rate much below the average growth rate and tries to run a Ponzi scheme by issuing a lot of debt and rolling it over forever, it is gambling. The policy may well work, but it might not. And the circumstances in which it fails are particularly dire. The big losers are the generations alive when the scheme fails, which must endure either a debt default or higher taxes. The failure is especially painful because it occurs in a state of the world with extraordinarily low growth and thus high marginal utility of consumption. A government running a Ponzi scheme with debt is like a homeowner canceling his fire insurance to save money or an investor selling deep out-of-the-money puts: it works most of the time, but when it doesn't, all hell breaks loose.[11]

Fourth, even if the economy is dynamically efficient in the sense of not accumulating excessive capital, there still might be some potential welfare improvements from intergenerational risk sharing.[12] From the perspective of time zero, a yet-to-be-born generation does not know whether it will arrive during a lucky or unlucky time, and it may want to share that risk with other generations. This intergenerational risk sharing can be achieved with well-designed fiscal policy. How this risk sharing interacts with debt policy is, I admit, still not completely clear to me, though some recent work explores this topic.[13] There are likely more papers to be written on this issue before it is resolved.

III. Adding Market Power

In addition to risk, another reason the interest rate on government debt can fall below the net marginal product of capital is market power. If firms charge prices above marginal cost, there is a wedge between the cost of capital (as reflected by market interest rates) and the marginal product of capital. The logic is straightforward. In the presence of market power, the price of output is a markup over marginal cost:

$$P = \mu \ MC.$$

11. See Mian, Straub, and Sufi (2022) for a recent contribution to the literature on debt sustainability. Their proposition 5 suggests that the increase in debt must be sufficiently small to guarantee success of the Ponzi scheme.

12. See Ball and Mankiw (2007) for one approach to this topic.

13. See, for example, Brumm and others (2021).

One measure of marginal cost is the cost of capital divided by the marginal product of capital:

$$MC = \frac{(r + \delta) P}{MPK}.$$

These two equations imply that the real interest rate is

$$r = \frac{MPK}{\mu} - \delta.$$

Thus, even under certainty, market power causes the real interest rate to fall below the net marginal product of capital. In a recent paper, Larry Ball and I calibrate this effect and conclude that the wedge is about 4 percentage points.[14]

The earlier equation for the steady-state real interest rate in the Solow model can be generalized for an economy with market power:

$$r = \alpha\left(\frac{n + g + \delta}{\mu s}\right) - \delta.$$

The markup attenuates the effects of saving and growth on the real interest rate (for a given α).[15] But this generalization also provides another reason that interest rates might have declined. Many observers have suggested that, over the past several decades, markets have become less competitive, and markups have increased.[16] Other things being equal, a higher markup reduces the equilibrium interest rate.[17]

14. Ball and Mankiw (2021) develop and calibrate a version of the Solow model that includes firms with market power and, because of fixed costs, increasing returns to scale. That paper shows that, in the presence of market power, the marginal product of capital can either exceed or fall short of measured capital income per unit of capital. In the realistic calibration presented there, the marginal product of capital exceeds capital income per unit of capital. This finding tends to reinforce the conclusion that the economy is dynamically efficient.

15. A nettlesome but important detail: calibrating α, the exponent on capital in the Cobb-Douglas production function, is now more difficult. In the competitive economy of the standard Solow model, α equals capital's share of income. That is not necessarily the case in an economy with market power. The calibration in Ball and Mankiw (2021) suggests that α is larger than the measured capital share. As a result, the effect of greater saving and lower growth on the real interest rate is only slightly smaller than in my earlier calculations for a competitive economy.

16. See, for example, Barkai (2020), De Loecker, Eeckhout, and Unger (2020), and Philippon (2019). The size of the change in markups is controversial; see Basu (2019).

17. Eggertsson, Robbins, and Wold (2018) explore this issue.

Again, we can get a sense of how large this effect might be. The previous equation implies

$$\frac{\partial r}{\partial \mu} = -\alpha\left(\frac{n+g+\delta}{\mu^2 s}\right).$$

With my calibrated parameters and a markup of, say, 20 percent (so $\mu = 1.2$), this becomes

$$\frac{\partial r}{\partial \mu} \approx -\left(\frac{1}{3}\right)\left(\frac{.01+.02+.05}{1.2^2(.24)}\right) = -.08.$$

An increase in the markup of 1 percentage point reduces the real interest rate by 8 basis points. Some of the literature suggests that markups have increased by 20 percentage points or more. This change could explain a decline in real interest rates of about 160 basis points.

The wedge induced by market power can have profound implications for fiscal policy. In a recent paper, Ball and I (2021) show that by reducing the interest rate, the wedge makes it easier for the government to roll over debt forever. But unlike in a competitive economy, a successful Ponzi scheme in an economy with market power can reduce welfare. When the government debt crowds out capital, the output loss from the smaller capital stock is determined not by the real interest rate but by the much higher marginal product of capital. Even if high government debt is benign from the standpoint of the budgetary sustainability, it can still reduce steady-state labor productivity, real wages, and aggregate consumption.

IV. Key Takeaways

So where does that leave us? Let me suggest four tentative conclusions.

First, the decline in real interest rates around the world over the past several decades is not a mystery. It appears to be the result of an increase in world saving, a decline in world growth, and possibly an increase in market power.

Second, because interest rates are so low, greater government debt is most likely not problematic from a budgetary standpoint. The government can probably roll over the debt and the accumulating interest forever, in essence letting growth take care of the debt.

Third, there is an outside chance that this Ponzi scheme of perpetual debt rollover will fail. That possible outcome is especially dire because the failure makes an already bad state of the world even worse.

Finally, even if the perpetual debt rollover succeeds, the increased debt could still crowd out capital. If the economy's capital stock is less than the golden rule level, as appears to be the case, this reduction in capital accumulation will, in the long run, depress not only labor productivity and real wages but also the resources available for consumption.

ACKNOWLEDGMENTS I am grateful to Laurence Ball, Olivier Blanchard, Rohit Goyal, Laurence Kotlikoff, Ludwig Straub, and Lawrence Summers for helpful discussion and comments.

References

Abel, Andrew B., N. Gregory Mankiw, Lawrence H. Summers, and Richard J. Zeckhauser. 1989. "Assessing Dynamic Efficiency: Theory and Evidence." *Review of Economic Studies* 56, no. 1: 1–19.

Ball, Laurence, Douglas W. Elmendorf, and N. Gregory Mankiw. 1998. "The Deficit Gamble." *Journal of Money, Credit and Banking* 30, no. 4: 699–720.

Ball, Laurence, and N. Gregory Mankiw. 2007. "Intergenerational Risk Sharing in the Spirit of Arrow, Debreu, and Rawls, with Applications to Social Security Design." *Journal of Political Economy* 115, no. 4: 523–47.

Ball, Laurence, and N. Gregory Mankiw. 2021. "Market Power in Neoclassical Growth Models." Working Paper 28538. Cambridge, Mass.: National Bureau of Economic Research, forthcoming in *Review of Economic Studies*. https://www.nber.org/papers/w28538.

Barkai, Simcha. 2020. "Declining Labor and Capital Shares." *Journal of Finance* 75, no. 5: 2421–63.

Basu, Susanto. 2019. "Are Price-Cost Markups Rising in the United States? A Discussion of the Evidence." *Journal of Economic Perspectives* 33, no. 3: 3–22.

Bernanke, Ben. 2005. "The Global Saving Glut and the US Current Account Deficit." Remarks at the Sandridge Lecture, Virginia Association of Economists, Richmond, Virginia, March 10. https://www.federalreserve.gov/boarddocs/speeches/2005/200503102/.

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

Brumm, Johannes, Xiangyu Feng, Laurence J. Kotlikoff, and Felix Kubler. 2021. "When Interest Rates Go Low, Should Public Debt Go High?" Working Paper 28951. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28951.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics* 135, no. 2: 561–644.

Diamond, Peter A. 1965. "National Debt in a Neoclassical Growth Model." *American Economic Review* 55, no. 5: 1126–50.

Eggertsson, Gauti B., Jacob A. Robbins, and Ella Getz Wold. 2018. "Kaldor and Piketty's Facts: The Rise of Monopoly Power in the United States." Working Paper 24287. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24287.

Mian, Atif, Ludwig Straub, and Amir Sufi. 2022. "A Goldilocks Theory of Fiscal Deficits." Working Paper 29707. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29707.

O'Connell, Stephen, and Stephen Zeldes. 1988. "Rational Ponzi Games." *International Economic Review* 29, no. 3: 431–50.

Phelps, Edmund. 1961. "The Golden Rule of Accumulation: A Fable for Growthmen." *American Economic Review* 51, no. 4: 638–43.

Philippon, Thomas. 2019. *The Great Reversal: How America Gave Up on Free Markets*. Cambridge, Mass.: Harvard University Press.

Rachel, Lukasz, and Thomas D. Smith. 2017. "Are Low Real Interest Rates Here to Stay?" *International Journal of Central Banking* 13, no. 3: 1–42.

Romer, David. 2019. *Advanced Macroeconomics*. 5th ed. New York: McGraw-Hill.

Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70, no. 1: 65–94.

Straub, Ludwig. 2019. "Consumption, Savings, and the Distribution of Permanent Income." Working Paper. https://scholar.harvard.edu/straub/publications/consumption-savings-and-distribution-permanent-income.

Summers, Lawrence H. 2016. "The Age of Secular Stagnation." *Foreign Affairs*, March/April. https://www.foreignaffairs.com/articles/united-states/2016-02-15/age-secular-stagnation.

PHILLIP L. SWAGEL
Congressional Budget Office

Budgetary Implications of Economic Scenarios with Higher and Lower Interest Rates

ABSTRACT This paper illustrates how the Congressional Budget Office's (CBO's) July 2021 baseline budget projections would have differed if the agency had used two alternative economic scenarios. The high-sixth scenario is based on the average values of projections for several variables—including inflation and the growth of gross domestic product after removing the effects of inflation (real GDP)—from the six *Blue Chip* forecasters (about one-sixth of the total) with the highest average interest rate projections. The low-sixth scenario is based on the average values of projections for the same variables from the six *Blue Chip* forecasters with the lowest average interest rate projections. Using its simplified model of how macroeconomic changes would affect the federal budget, the CBO found that projected deficits would be $2.1 trillion larger from 2022 to 2031 under the high-sixth scenario (totaling $13.8 trillion) than under the low-sixth scenario ($11.7 trillion). Despite a greater amount of debt in dollar terms under the high-sixth scenario, federal debt held by the public as a percentage of GDP would total about 101 percent at the end of 2031 under both scenarios.

Alternative forecasts of economic variables can have large effects on the Congressional Budget Office's (CBO's) projections of the budget. For its source of alternative forecasts in this analysis, the CBO used those from private-sector economists as reported in *Blue Chip Economic Indicators* in

Conflict of Interest Disclosure: Phillip L. Swagel is the director of the Congressional Budget Office (CBO). 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. Other than the aforementioned, he is currently not an officer, director, or board member of any organization with an interest in this paper.

Brookings Papers on Economic Activity, Spring 2022: 233–249

March 2022.[1] The CBO regularly shows its forecast in relation to the middle two-thirds of the range of the *Blue Chip* forecasts; that range reflects some of the uncertainty about the economic projections.

To produce alternative budget estimates starting with the data from the *Blue Chip* forecasters, the CBO added projections of macroeconomic variables from a statistical model and then used a simplified budget model. This paper reports the CBO's projections of economic and budgetary outcomes under two scenarios with divergent paths for interest rates.

I. Data the CBO Used from *Blue Chip* Forecasters

For calendar years 2022 and 2023, the CBO used values of projections for eight variables made by the thirty-eight *Blue Chip* forecasters whose complete interest rate projections were reported: the three-month Treasury bill rate, the ten-year Treasury note rate, the unemployment rate, growth of real GDP (adjusted to remove the effects of inflation), growth of real personal consumption expenditures, growth of real nonresidential fixed investment, inflation as measured by the GDP price index, and inflation as measured by the consumer price index for all urban consumers (CPI-U).

To determine the highest and lowest interest rate projections, the CBO calculated the average of the projections for three-month and ten-year Treasury rates and then sorted those averages first by the values for 2022 and then by the values for 2023. For each economic variable, the 2022 value in the high-sixth scenario was the average among the six forecasters with the highest interest rate projections. Similarly, the 2022 value in the low-sixth scenario was the average among the six forecasters with the lowest interest rate projections. The 2023 values under both scenarios were calculated in the same way for that year.

For projections of interest rates and inflation for 2024 to 2031, *Blue Chip Economic Indicators* reports averages for the top ten and bottom ten forecasters rather than projections for each forecaster and does not report which forecasters are included in those averages. (The rank order of forecasters changes from year to year.) The CBO used the information on the average for the highest ten and the lowest ten interest rate and inflation projections by *Blue Chip* forecasters for the high-sixth and low-sixth scenarios, respectively. For interest rates, the CBO used the three-month Treasury bill rate and the ten-year Treasury note rate. For inflation, the CBO used measures

1. Wolters Kluwer, "Blue Chip Publications," https://www.wolterskluwer.com/en/solutions/vitallaw-law-firms/blue-chip.

based on the GDP price index and the CPI-U. Thus, for 2024, for example, the average of the highest ten projections only roughly approximates the 2024 projections of interest rates by the six forecasters whose forecasts were used for 2022 and 2023 in the high-sixth scenario. For the highest ten projections, *Blue Chip Economic Indicators* reports a single average value for 2029 to 2033, and the CBO used that value for each year from 2029 to 2031 in the high-sixth scenario. For the lowest ten projections, the CBO used the analogous value in the low-sixth scenario.

For projections of real GDP growth rates and unemployment rates for 2024 to 2031, the CBO used information from the *Blue Chip* consensus (the average of all *Blue Chip* forecasters) for both the high-sixth and low-sixth scenarios. The CBO used those rates because the rates for 2022 and 2023 for both scenarios were similar to the *Blue Chip* consensus in those years, and the rates for the specific forecasters used in the scenarios for 2023 were not available for later years. For the consensus, *Blue Chip Economic Indicators* reports a single average value for 2029 to 2033, and the CBO used that value for each year from 2029 to 2031.

II. How the CBO Projected Additional Macroeconomic Variables from a Statistical Model

To obtain the full set of economic variables needed for budget estimates, the CBO augmented data from the *Blue Chip* forecasters with projections from a statistical model that draws on historical correlations between macroeconomic variables to produce conditional forecasts. As targets for that model, the agency used data on actual values for 2021 when available and values from the *Blue Chip* forecasters described in the previous section.

The CBO projected values for 2022 to 2031 for ten additional macroeconomic variables that are used in its simplified budget model: the federal funds rate (the rate that financial institutions charge each other for overnight loans of their monetary reserves), payroll employment, the number of people in the labor force, wages and salaries, nominal gross national product (GNP), nominal private nonresidential fixed investment in equipment, nominal potential GDP (the maximum sustainable output of the economy), real potential GDP, inflation as measured by the consumer price index for food at home, and inflation as measured by the consumer price index for medical care.

In addition, the CBO projected values for 2024 to 2031 for the growth of real personal consumption expenditures and the growth of real nonresidential fixed investment. The projections of those variables for 2022

and 2023 were derived from the *Blue Chip* forecasts as discussed in the previous section.

The CBO adapted its approach to conditional forecasting from that used by the staff of the Federal Reserve Bank of New York, who have used a Bayesian vector autoregression (BVAR) and the Kalman filter (or smoother) for a similar purpose.[2] Bayesian techniques are particularly well suited to estimating parameters in a large system of equations given a limited amount of data. The modeling is structured, for example, so that a projection of a variable at a given point in time is more likely to be influenced by recent data than by older data. The structure prevents the estimation from explaining historical data very well while having a poor ability to forecast beyond the data used for estimation—which would be the case if the estimation process overfit the parameters. The approach is very flexible, and the staff of the Federal Reserve Bank of New York found that it generated reasonable conditional forecasts. The Kalman filter fills out the projection of all the variables over the entire forecast period around a projection of any subset of those variables in any period—essentially starting with the *Blue Chip* forecasts discussed above and imputing forecasts of additional variables.

III. How the CBO Estimated the Effects of Macroeconomic Changes on the Federal Budget

Using its budgetary feedback model (BFM) and some baseline projection methods, the CBO analyzed the effects of macroeconomic changes on the federal budget under the high-sixth and low-sixth scenarios relative to its July 2021 baseline.[3]

The eighteen variables described in the previous two sections are used by a component of the BFM that generates projections of related variables used in the budget projections. For example, those variables include income as measured in the national income and product accounts (NIPAs), the employment cost index, and interest rates on Treasury securities over a full range of maturities (determined using the federal funds rate, the three-month Treasury bill rate, and the ten-year Treasury note rate as a basis).

2. See Crump and others (2021). Using the approach outlined in that paper, the CBO conducted its Bayesian analysis using the Minnesota prior, in which economic variables follow independent random-walk processes potentially with drift. The factors setting the tightness of that prior were estimated on the basis of the amount of data relative to the number of parameters in the model.

3. At the time this paper was written, the CBO had not yet published updated economic and budget projections in 2022; see CBO (2022) for that update.

The BFM provides a unified framework to approximate the budgetary feedback that would be arrived at by using a wider array of the CBO's budgetary models.[4] Because the agency's baseline projection method better captures the budgetary effects arising from larger changes in interest rates and inflation than the BFM does, the agency used its baseline projection method to estimate some of those costs.

III.A. Revenues

To estimate effects on revenues, the CBO generally estimates changes to macroeconomic variables as measured in the NIPAs. The relationship between changes in certain NIPA measures and changes in revenues may be straightforward. For example, as wages and salaries, corporate profits, and imports increase, so do the receipts from income and payroll taxes and customs duties. The BFM applies estimated marginal tax rates appropriate for NIPA measures, which differ from the effective marginal tax rates on corresponding taxable income used in the CBO's specialized models. When a component of taxable income is not directly measured in the NIPAs, one or more alternative NIPA measures are used as proxies. For example, the BFM derives an estimate of the change in profits from S corporations from a combination of the NIPA measures for proprietors' income, domestic corporate profits, and estimates from the CBO's full modeling of the share of domestic corporate profits that represents S corporations' profits. Certain taxes, such as those on capital gains or inherited property, are sensitive to changes in the stock market and other asset values, but those macroeconomic changes are not directly measured in NIPA accounting. The BFM uses changes in nominal GDP to approximate how macroeconomic changes would affect revenues from the estate tax and capital gains.

III.B. Spending

To estimate effects on federal spending, the CBO considered how the three broad categories of outlays in the budget—discretionary, mandatory, and net interest—would respond to changes in macroeconomic variables. Discretionary spending is calculated according to statutory requirements using a weighted average of the employment cost index and the GDP deflator. (For those calculations, the CBO used its baseline projection methods.)

4. For a description of how the BFM is constructed, what parameter values it incorporates, how the CBO uses the model, and its limitations, see Frentz and others (2020). For a simpler interactive tool based on the BFM, see CBO, "Workbook for How Changes in Economic Conditions Might Affect the Federal Budget, June 2021," https://www.cbo.gov/publication/57191.

Mandatory spending consists of federal noninterest spending that is not generally governed by the annual appropriation process. Outlays for mandatory programs usually depend on the number of program participants and the level of spending per participant, which is often tied directly to changes in the macroeconomy. The estimates reflect changes in eight components of the budget: Social Security, other indexed entitlements, Medicare, Medicaid, unemployment compensation, the Supplemental Nutrition Assistance Program (SNAP), refundable tax credits, and child nutrition programs. (To calculate Social Security spending, the CBO used its baseline projection methods.) Although changes in the macroeconomy could affect other programs, the BFM does not include those effects because the CBO estimates they are approximately zero, on average.

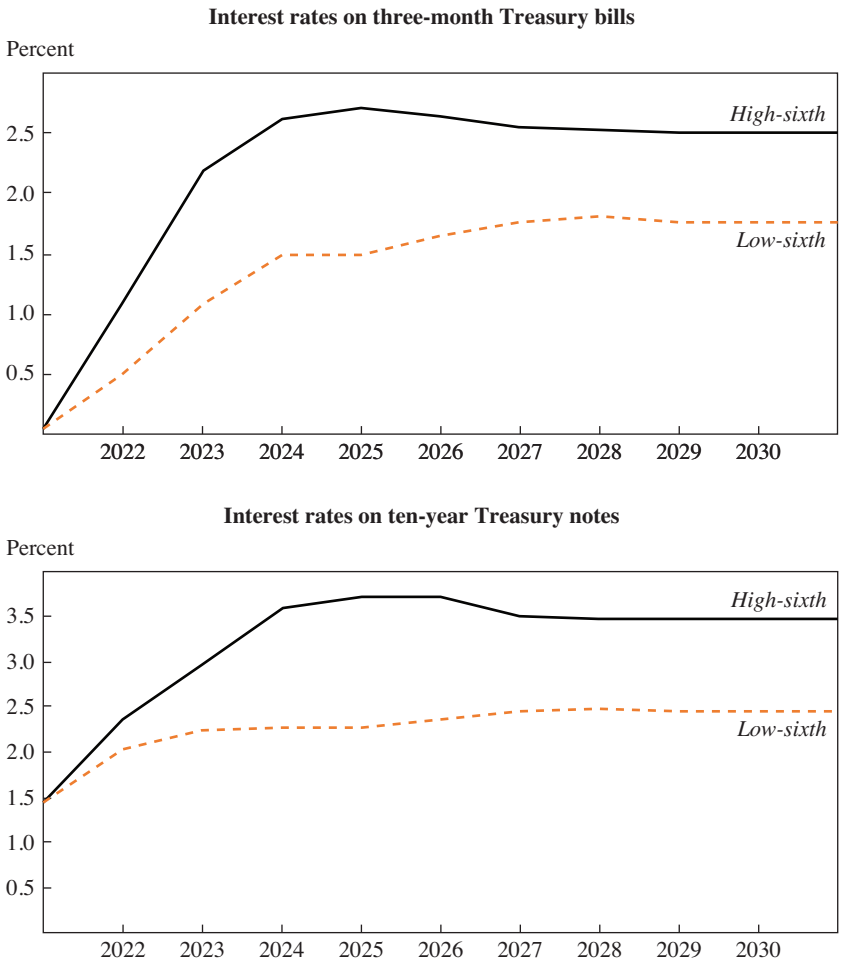
To estimate spending for net interest in this analysis, the CBO used its baseline projection method, which better captures the effects of larger changes in interest rates than the BFM does. The stock of federal debt at the beginning of a projection period, along with the additional debt (generally the amount of annual projected deficits), substantially determines the amount of outstanding debt in each year. In addition, the CBO projected the characteristics of debt that the Treasury would issue to finance annual deficits—for example, the time to maturity, whether interest rates are fixed or floating, and whether the interest payments include an adjustment for inflation. The CBO used estimates of interest rates related to those characteristics to determine the amounts of interest that the Treasury would pay on outstanding debt.

IV. Economic Projections under the Scenarios

The high-sixth and low-sixth scenarios group the forecasts with the highest and lowest projections of interest rates (see figure 1). For example, these are the forecasts made in March 2022 for the calendar year average of interest rates on three-month Treasury bills: the highest one-sixth of *Blue Chip* forecasts, taken together, averages 1.1 percent in 2022 and 2.2 percent in 2023; and the lowest one-sixth of *Blue Chip* forecasts, taken together, averages 0.5 percent in 2022 and 1.1 percent in 2023.

The CBO does not interpret the differences in average interest rates between the high-sixth and low-sixth scenarios as representing possible forecast errors of interest rates. One measure of such possible errors is the two-thirds spread—that is, the difference between the 17th percentile and the 83rd percentile—of the CBO's errors over two years. In the CBO's

Figure 1. Interest Rates



Sources: Congressional Budget Office; Federal Reserve; Wolters Kluwer, *Blue Chip Economic Indicators*, vol. 47, no. 3 (March 11, 2022).

Note: The CBO constructed the high-sixth and low-sixth scenarios based on the average values of projections from the six *Blue Chip* forecasters (about one-sixth of the total) with the highest and lowest average interest rate projections, respectively, for 2022 and 2023.

Table 1. Economic Projections

Percent	*2022*	*2023*	*2024*	*2025*	*Total 2026–2031*
Interest rate on three-month Treasury bills					
High-sixth scenario	1.1	2.2	2.6	2.7	2.5
Low-sixth scenario	0.5	1.1	1.5	1.5	1.8
Interest rate on ten-year Treasury notes					
High-sixth scenario	2.4	3.0	3.6	3.7	3.5
Low-sixth scenario	2.0	2.3	2.3	2.3	2.4
Inflation as measured by the GDP price index					
High-sixth scenario	5.2	3.3	2.6	2.5	2.3
Low-sixth scenario	4.7	2.3	1.9	1.8	1.8
Real GDP growth					
High-sixth scenario	3.7	2.6	2.1	2.0	1.9
Low-sixth scenario	3.2	2.4	2.1	2.0	1.9
Unemployment rate					
High-sixth scenario	3.6	3.3	3.6	3.7	3.9
Low-sixth scenario	3.7	3.7	3.6	3.7	3.9

Sources: Congressional Budget Office; Wolters Kluwer, *Blue Chip Economic Indicators*, vol. 47, no. 3 (March 11, 2022).

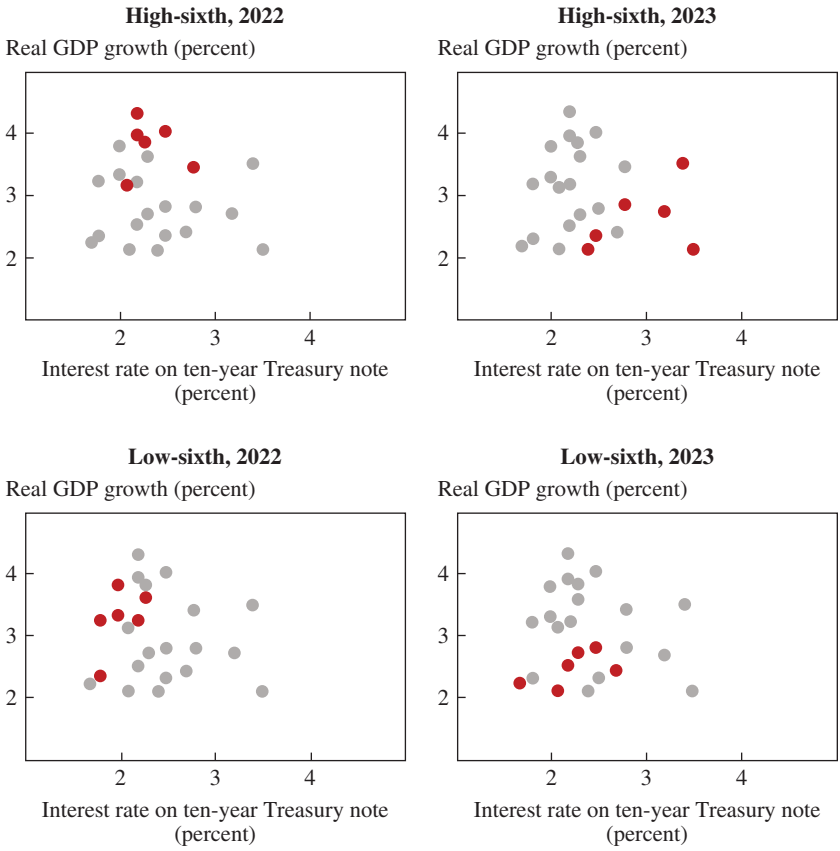
Note: Data are annual averages for calendar years. The CBO constructed the high-sixth and low-sixth scenarios based on the average values of projections from the six *Blue Chip* forecasters (about one-sixth of the total) with the highest and lowest average interest rate projections, respectively, for 2022 and 2023.

forecasts of the three-month Treasury bill rate from 1981 to 2019, that spread was 2.3 percentage points (CBO 2021b).

Rather than illustrating how large the forecast errors might be, the range of *Blue Chip* forecasts illustrates how different the CBO's forecast might have been. Historically, the CBO's forecasts have been within the middle two-thirds of the range of *Blue Chip* forecasts for most variables in most years. The CBO generally aims for its forecasts to be in the middle of a range of likely outcomes, and large forecast errors arise when outcomes are far from that middle.

In contrast to the differences between the scenarios' projections of interest rates, the projections of unemployment rates and real GDP growth rates for 2022 and 2023 are similar in the two scenarios, in the CBO's assessment (see table 1). For example, the projection of the unemployment rate for 2022 is 3.6 percent in the high-sixth scenario and 3.7 percent in the low-sixth scenario. That similarity indicates that the differences in interest rate forecasts are not simply the result of differences in the projected strength of the economy across the board. Taken individually, projections from the *Blue Chip* forecasters that contributed to each scenario show a variety of relationships between interest rates and real GDP growth for 2022 and also

Figure 2. Real GDP Growth and Interest Rates



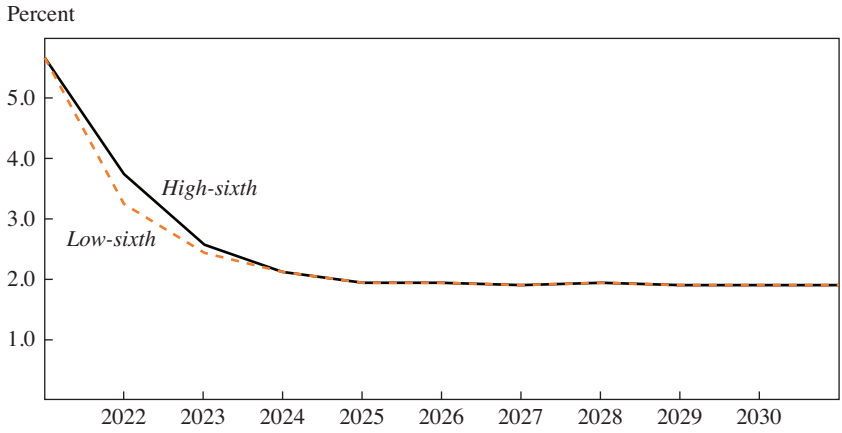
Sources: Congressional Budget Office; Bureau of Economic Analysis; Federal Reserve; Wolters Kluwer, *Blue Chip Economic Indicators,* vol. 47, no. 3 (March 11, 2022).

Note: The data points represent forecasts by each *Blue Chip* forecaster. The CBO constructed the high-sixth and low-sixth scenarios based on the average values of projections from the six *Blue Chip* forecasters (about one-sixth of the total) with the highest and lowest average interest rate projections, respectively, for 2022 and 2023.

for 2023 (see figure 2). Taken together, the forecasts that constitute the high-sixth scenario project slightly faster GDP growth for 2022 and 2023 than those that make up the low-sixth scenario (see figure 3).

By constructing the scenarios using the same real GDP growth rates and unemployment rates for 2024 to 2031, the CBO was able to focus the analysis on the differences between the scenarios in interest rates and the associated rates of inflation—which also are higher in the high-sixth scenario than in the low-sixth scenario (see figure 4).

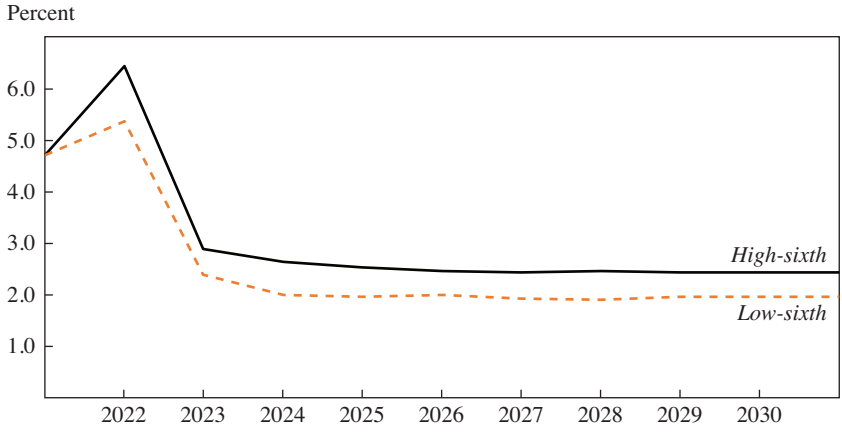
Figure 3. Real GDP Growth



Sources: Congressional Budget Office; Bureau of Economic Analysis; Wolters Kluwer, *Blue Chip Economic Indicators,* vol. 47, no. 3 (March 11, 2022).

Note: The CBO constructed the high-sixth and low-sixth scenarios based on the average values of projections from the six *Blue Chip* forecasters (about one-sixth of the total) with the highest and lowest average interest rate projections, respectively, for 2022 and 2023.

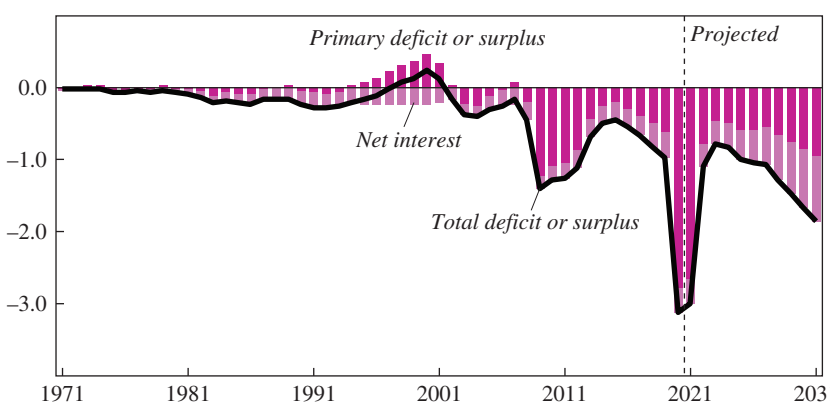
Figure 4. Inflation as Measured by the Consumer Price Index for All Urban Consumers



Sources: Congressional Budget Office; Bureau of Labor Statistics; Wolters Kluwer, *Blue Chip Economic Indicators,* vol. 47, no. 3 (March 11, 2022).

Note: The CBO constructed the high-sixth and low-sixth scenarios based on the average values of projections from the six *Blue Chip* forecasters (about one-sixth of the total) with the highest and lowest average interest rate projections, respectively, for 2022 and 2023.

Figure 5. Deficits in CBO's July 2021 Baseline Budget Projections



Source: Congressional Budget Office.

V. Deficit Projections under the Scenarios

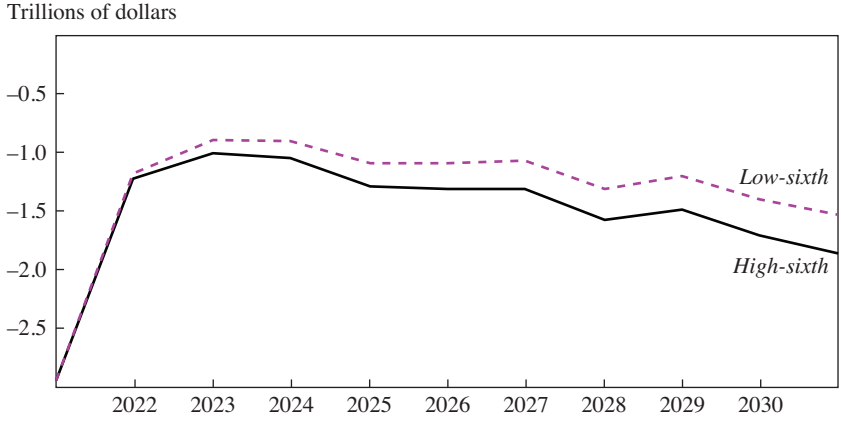
In the CBO's July 2021 baseline budget projections, total deficits were projected to grow substantially over the 2022–2031 period (CBO 2021a). Much of that growth stemmed from projected increases in spending for net interest (see figure 5).

If the CBO had used the economic projections in the high-sixth scenario to underlie its July 2021 budget projections, the deficit would be larger in every year of the projection period than if the agency had used the projections in the low-sixth scenario (see figure 6 and table 2). Over the ten-year period, total deficits would be $2.1 trillion larger under the high-sixth scenario than under the low-sixth scenario, the CBO projects.

Those larger deficits are the result of effects on revenues and mandatory spending that largely offset each other and effects on net interest costs. Under the high-sixth scenario, revenues increase more mainly because of higher taxable income, and noninterest spending increases more primarily because of cost-of-living adjustments for benefit programs and higher prices for the purchase of goods and services by the federal government.

Net interest costs are higher under the high-sixth scenario mostly because of the higher interest rates on debt issued by the Treasury. One way to assess the burden to the government of net interest is to measure the cost as a percentage of GDP—2.9 percent in the high-sixth scenario and 2.1 percent in the low-sixth scenario in 2031 (see figure 7).

Figure 6. Deficits under the High-Sixth and Low-Sixth Scenarios



Source: Congressional Budget Office.

Note: CBO constructed the high-sixth and low-sixth scenarios based on the average values of projections from the six *Blue Chip* forecasters (about one-sixth of the total) with the highest and lowest average interest rate projections, respectively, for 2022 and 2023.

Table 2. Cumulative Budgetary Effects of the High-Sixth Scenario Relative to the Low-Sixth Scenario from 2022 to 2031

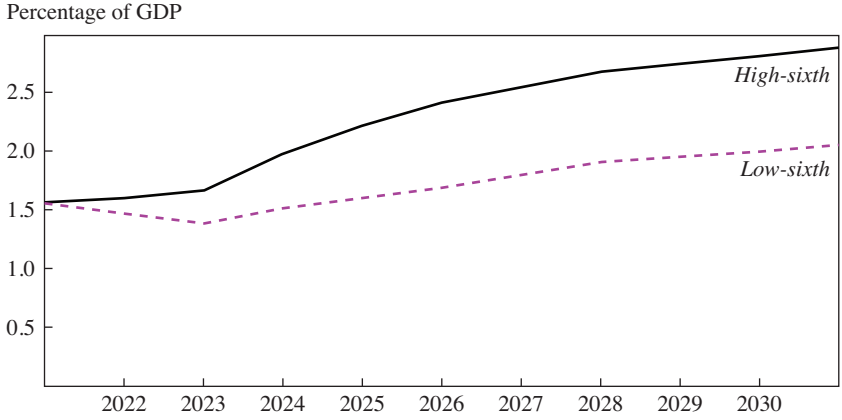
Trillions of dollars	*Increase (–) or decrease in the deficit*
Revenue effects	1.6
Noninterest spending effects	–1.5
Net interest spending effects	–2.2
Total	**–2.1**

Source: Congressional Budget Office.

The budgetary implications of the two scenarios are uncertain. For example, because changes in revenues and noninterest outlays roughly offset each other for a given change in inflation in this analysis, primary deficits are not very sensitive to changes in inflation. However, the change in primary deficits could be larger or smaller than the CBO's estimates for a given change in inflation—depending, for instance, on the types of goods and services for which inflation is strongest and how taxable income may change as a result.

Higher interest rates projected by *Blue Chip* forecasters and the associated higher inflation projected by those forecasters are the two main factors leading to the larger deficits, on net, in the high-sixth scenario than in

Figure 7. Net Interest Spending



Source: Congressional Budget Office.

Note: The CBO constructed the high-sixth and low-sixth scenarios based on the average values of projections from the six *Blue Chip* forecasters (about one-sixth of the total) with the highest and lowest average interest rate projections, respectively, for 2022 and 2023.

the low-sixth scenario. The remainder of this section describes the ways that higher interest rates and inflation affect the CBO's budget estimates.

V.A. Revenues

Larger increases in wage rates and prices generally lead to greater labor income, profits, and other nominal income, which in turn generate larger collections of individual income taxes, payroll taxes, and corporate income taxes. Revenues from the Federal Reserve's remittances to the Treasury are affected by changes in interest rates.

Many provisions in the individual income tax system—including the income thresholds for the tax brackets—are adjusted, or indexed, for inflation. Therefore, the share of taxpayers' income that is taxed at certain rates does not change very much when income increases because of higher inflation, so tax collections tend to rise roughly proportionally with income under those circumstances. However, not all parameters of the individual income tax system are indexed for inflation. For example, the income thresholds for the surtax on investment income are fixed in nominal dollars, so if income rose because of inflation, the surtax would apply to a larger share of taxpayers' income.

For the payroll tax, rates mostly stay the same when earnings change, with two main exceptions—when earnings reach the maximum amount

subject to the Social Security tax (although that maximum amount rises, after a lag, with average wages in the economy) and when they reach the threshold for the additional Medicare tax. Higher wage growth thus leads to a roughly proportional increase in payroll tax revenues, the CBO estimates.

Similarly, nearly all corporate profits are taxed at a single statutory rate of 21 percent. Consequently, an increase in profits resulting from higher inflation generates a roughly proportional increase in corporate tax revenues.

Finally, higher nominal interest rates would first reduce and then increase revenues from the Federal Reserve's remittances to the Treasury. If, over the next ten years, all interest rates were higher than the CBO projects, the Federal Reserve's remittances over the next few years would be smaller than projected because higher interest payments on deposits that banks hold at the central bank would outstrip the additional earnings from interest on its Treasury holdings. Over time, however, the current holdings in the portfolio would mature and be replaced with higher-yielding investments; as a result, the Federal Reserve's remittances would be larger in later years.

V.B. Mandatory Spending

Higher inflation would increase the cost of several mandatory spending programs, the CBO estimates. Benefits for many mandatory programs are automatically adjusted each year to reflect increases in prices. Specifically, benefits paid for Social Security, federal employees' retirement programs, disability compensation for veterans, SNAP, Supplemental Security Income, child nutrition programs, and the refundable portion of the earned income tax credit, among others, are adjusted (with a lag) for changes in the consumer price index, one of its components, or another measure of inflation.

Many of Medicare's payment rates are also adjusted annually for inflation. Spending for some other programs, such as Medicaid, is not formally indexed to changes in prices but nevertheless tends to grow when inflation rises because the costs of providing benefits under those programs increase as wages and prices rise. In addition, to the extent that benefit payments in retirement and disability programs are linked to participants' preenrollment wages, higher wages would boost future outlays for those programs.

V.C. Discretionary Spending

As specified in law, the CBO's projections of spending for discretionary programs reflect the assumption that discretionary budget authority provided in appropriation acts in future years will match current funding, with adjustments for inflation. Hence, higher inflation would increase the CBO's

Table 3. Debt and GDP Projections

	High-sixth scenario	*Low-sixth scenario*	*CBO's July 2021 baseline*
	Debt as a percentage of GDP		
Debt in 2021	103	103	103
Debt in 2026	97	98	100
Debt in 2031	101	101	106
	Debt and GDP growth		
Growth in debt from 2021 to 2031 (%)	63	54	56
Growth in GDP from 2021 to 2031 (%)	66	56	50
Ratio of the debt growth rate to the GDP growth rate	0.96	0.96	1.11

Source: Congressional Budget Office.

Note: Data are for fiscal years. Debt is measured as debt held by the public. The CBO constructed the high-sixth and low-sixth scenarios based on the average values of projections from the six *Blue Chip* forecasters (about one-sixth of the total) with the highest and lowest average interest rate projections, respectively, for 2022 and 2023.

projections of such spending. The agency estimated the increase by adjusting discretionary budget authority to account for changes in wage and price indexes. Congressional decisions about annual appropriations would ultimately determine spending for discretionary programs.

V.D. Net Interest Costs

Higher inflation and interest rates would boost net interest spending in dollar terms. As the Treasury replaced maturing securities and increased its borrowing to cover deficits, the budgetary effects of higher interest rates would mount. Higher inflation would also make outstanding inflation-indexed securities more costly for the federal government. Moreover, greater deficits would increase federal debt, boosting interest costs further.

VI. Debt Projections under the Scenarios

Debt that the Treasury issues to the public is larger in the high-sixth scenario than in the low-sixth scenario owing to the larger budget deficits that stem mainly from higher net interest costs. Despite a greater amount of debt in dollar terms, federal debt held by the public as a percentage of GDP would be about the same at the end of 2031 under the two scenarios. That outcome would occur because the ratio of the growth rate of debt to the growth rate of nominal GDP would be roughly the same in the two scenarios (see table 3).

The higher inflation in the high-sixth scenario affects not only the amount of debt but also how burdensome that debt is. When inflation exceeds the rates expected when debt is issued at a fixed interest rate, some of the value of that debt (and its purchasing power) is transferred from lenders to borrowers. As a result, borrowers can spend a smaller share of their income repaying money they borrowed in the past when income rises with inflation, as it does in this analysis. Similarly, the federal government, as a borrower, can use a smaller share of the revenues it collects to pay holders of Treasury securities that mature, without changing tax rates. In that sense, the debt burden to the government is smaller when inflation is higher.

ACKNOWLEDGMENTS I thank members of the CBO's economic scenarios and budgetary implications team: Devrim Demirel, Nathaniel Frentz, Jeffrey Kling, Avi Lerner, Michael McGrane, Noah Meyerson, Jaeger Nelson, Dan Ready, John Seliski, Emily Stern, Jeffrey Werling, Matthew Wilson, and Byoung Hark Yoo. Christine Bogusz edited this paper, and Grace Berry checked the results; Casey Labrack and R. L. Rebach created the graphics. This work was managed by Christina Hawley Anthony, Robert Arnold, Mark Doms, Michael Falkenheim, Sebastien Gay, Theresa Gullo, John Kitchen, Leo Lex, John McClelland, Sam Papenfuss, Joseph Rosenberg, Joshua Shakin, and Robert Sunshine.

References

Congressional Budget Office (CBO). 2021a. *Additional Information about the Updated Budget and Economic Outlook: 2021 to 2031*. https://www.cbo.gov/publication/57373.

Congressional Budget Office (CBO). 2021b. *CBO's Economic Forecasting Record: 2021 Update*. https://www.cbo.gov/publication/57684.

Congressional Budget Office (CBO). 2022. *The Budget and Economic Outlook: 2022 to 2032*. https://www.cbo.gov/publication/57950.

Crump, Richard K., Stefano Eusepi, Domenico Giannone, Eric Qian, and Argia M. Sbordone. 2021. "A Large Bayesian VAR of the United States Economy." Staff Report 976. New York: Federal Reserve Bank of New York. http://newyorkfed.org/research/staff_reports/sr976.html.

Frentz, Nathaniel, Jaeger Nelson, Dan Ready, and John Seliski. 2020. "A Simplified Model of How Macroeconomic Changes Affect the Federal Budget." Working Paper 2020-01. Washington: Congressional Budget Office. https://www.cbo.gov/publication/55884.

General Discussion

GENERAL DISCUSSION Olivier Blanchard commented that he agrees with Gregory Mankiw on the uncertainty of whether there are dynamic efficiencies or inefficiencies—there are many distortions that matter, but the welfare cost is probably lower than it was in the past. Blanchard argued that the effective lower bound was missing from the discussion and that it matters because it may come into play to sustain demand. Blanchard disagreed with Carmen Reinhart that the decrease in real rates is largely a result of monetary policy. The decrease started long before quantitative easing, the financial crisis, and COVID-19 and while quantitative easing allowed banks to achieve lower de facto interest rates, the causality comes from real factors, not policy. Quantitative easing is not monetization in the usual sense; it is a swap of two liabilities: bonds and interest-paying central bank reserves, he concluded. On Phillip Swagel's remarks, Blanchard noted that he suspects the scenarios presented were benign relative to what we may expect would happen; he went on to suggest that the scenarios should be ranked by the difference between the interest rate and inflation rate, which is what matters for debt dynamics.

Steven Davis remarked that safe government debt provides a convenience yield for many holders of collateral that supports other transactions. He pondered whether we should use the total return, inclusive of the convenience yield, when we evaluate the long-term forces that determine the safe real interest rate. He further posed the question whether convenience yields have risen enough in recent decades to account for much of the decline in pecuniary return on safe government debt.

Austan Goolsbee was struck by the view in Mankiw's presentation that suggests an increase in savings and a lack of investment opportunities as

Brookings Papers on Economic Activity, Spring 2022: 250–252

the cause of low interest rates. He argued that if one takes the view that the savings glut is caused by financial repression in China or other emerging markets, the risk of the saving rate going down and driving the interest rate up seems like a substantial risk worth considering as one scenario related to what will happen to the debt.

Mankiw responded to Davis and confirmed that the convenience yield is indeed left out of the neoclassical model; he indicated that it is not straightforward to calibrate. He agreed with Goolsbee that some of these forces can reverse but noted that the increase in the saving rate is likely the least important of the three—an increasing saving rate, lower growth, and higher markups. On Swagel's presentation, Mankiw said that he appreciated that there were multiple contingencies, multiple paths instead of just one. He was not convinced that figuring out the contingencies by looking at the tails of the *Blue Chip* forecasts is the best approach; instead, he suggested, by running an autoregression one could look at the standard errors and then find the outcome for the 10th percentile, the 20th percentile, and so on, and use those as scenarios. Mankiw concluded that such an approach would likely lead to a vastly more uncertain outlook for the future than the scenarios Swagel presented.

Reinhart argued that the big reserve accumulation of China has had a global impact on the savings glut and the lower interest rates. Reversal scenarios are not implausible and should be considered, for example, declining saving rates as demographic and housing problems in China accentuate. She speculated that the issue of the use of the dollar in sanctions could reduce the appetite for Treasuries over time. Reinhart moved on to address Blanchard's comment, saying that real factors do indeed matter, but she argued that Blanchard grossly underestimates the impact of monetary policy. The peak in real interest rates was in the early 1980s, she noted, following Paul Volcker raising the federal funds rate by almost 600 basis points. The cluster of exceptionally high real interest rates around that time was driven by monetary policy and inflation stabilization, marking the turning point for the secular decline. She asserted that she would not be convinced that monetary policy does not matter.

Blanchard clarified in the virtual conferencing chat that he agreed that central banks can affect the actual rate, something which was certainly the case with Volcker, but he argued that the decline since 1990 reflects a decrease in the real neutral rate.

Swagel noted that, in his comment, he showed two different scenarios, not two different economies, and he added that he and his colleagues will be working on more extreme scenarios in future research. In response to a

question in the chat, he clarified that the data he presented were publicly held debt.

Robert Hall emphasized the importance of the point made by Blanchard that when the government buys securities from the Federal Reserve, they are doing two things: funding the operation by borrowing and holding the corresponding securities. He pointed out that it is a matter of shifting between agencies and that there is no financial principle suggesting that it carries any importance beyond that. In addition, the Treasury tends to offset what the Federal Reserve does by switching to longer maturities when quantitative easing is in effect, he continued, revealing a huge coordination failure between the agencies but having no material effect on the capital market. He concluded that we do not need to worry about quantitative easing.

Chris Sims expressed uncertainty about any analysis that focuses on the period between 1980 and the present. He reiterated what Reinhart pointed out: the peak of real rates was in the early 1980s and the rise lasted about as long as the subsequent fall. He argued that it is not clear the rise can be explained by the same factors that some are invoking to understand the subsequent fall in rates.

Mankiw agreed with Reinhart that the period in the early 1980s was related to Paul Volcker's Federal Reserve policy and suggested that the decline up to about 1995 was at least partly related to monetary policy, but he contended that subsequent years were likely driven by real factors. He concluded by noting that what we are facing now—basically thirty-year Treasury inflation-protected securities at zero percent—is probably historically unprecedented.

JOHN FITZGERALD
Bowdoin College

ROBERT MOFFITT
Johns Hopkins University

The Supplemental Expenditure Poverty Measure: A New Method for Measuring Poverty

ABSTRACT We propose a new measure of the rate of poverty we call the supplemental expenditure poverty measure (SEPM), based on expenditure in the Consumer Expenditure Survey. It treats household expenditure as a measure of resources available to purchase the minimum bundle necessary to meet basic needs. Our measure differs from conventional income and consumption poverty in both concept and measurement, and it has advantages relative to both. Poverty rates using our basic measure are very close in level and recent trend to those of the most preferred income-based poverty rate produced by the US Census Bureau. But the SEPM poverty rate differs from the US Census Bureau measure at different levels of the poverty line. For example, the number of individuals living in either poor or almost poor households is 5 percentage points greater (about 16 million individuals) using our measure. We also construct an augmented measure that adds additional potential liquid resources. This "maximal resources" measure indicates that if disadvantaged households used up all their bank balances and maximized their credit card borrowing, 9.6 percent of the population (over 31 million individuals) would still be poor and unable to purchase the goods necessary for the basic needs of life.

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 officers, directors, or board members of any organization with an interest in this paper.

Brookings Papers on Economic Activity, Spring 2022: 253–286

The measurement of poverty has drawn the attention of economists for many decades. Both the level of poverty and its trend over time are important social indicators of the economic well-being of the most disadvantaged members of the society. Estimates of how poverty is affected by government policy in general, and by specific anti-poverty programs in particular, are also important indicators of the influence of government on improving the well-being of its poorest citizens. Nevertheless, how to best measure poverty has been the subject of significant disagreement among researchers and policy analysts.

There is renewed interest in the measurement of poverty in the United States. The US Census Bureau has recently conducted a major study of its most preferred poverty measure, the supplemental poverty measure (which the US Census Bureau abbreviates as SPM, and which we will designate SIPM for reasons given below), and how it could be improved. The study has recommended that the basic structure of the measure be retained but that a number of technical improvements be made.[1] A federal interagency working group established in 2019 and charged with studying alternative ways to measure poverty recently issued its report and recommended that an additional measure of poverty based on consumption rather than income be added to the measures produced by the US Census Bureau (OMB 2021). And the National Academies of Science, Engineering, and Medicine has formed an expert panel to spend two years studying additional improvements that might be made in the SIPM, with the panel slated to issue its final report in late 2022 or early 2023.

To supplement this activity, our study suggests a new method of measuring poverty that could be added to the two that have received the most attention in these discussions. The two methods are those that measure poverty by a household's income or its consumption. In both cases, the basic method is to start with some definition of the minimum bundle of goods that are needed to provide the basic needs of life. The minimum bundle is ultimately socially determined because what it means to be poor is a subjective concept that is up to the members of society to define. Starting with that minimum bundle, an income measure of poverty asks whether a household has enough income to purchase that bundle, while a consumption measure of poverty simply asks whether a household's level of consumption is sufficient to allocate enough consumption toward the goods in the

1. US Census Bureau, "Improvements to the Census Bureau's Supplemental Poverty Measure for 2021," https://www.census.gov/topics/income-poverty/supplemental-poverty-measure/library/working-papers/topics/potential-changes.html.

bundle to meet the minimum. In the language used in poverty measurement, both involve measuring a family's income (resources) or consumption to the threshold, which is the amount of income or consumption needed to meet the minimum bundle. A household is deemed poor if it does not have enough to meet that threshold and deemed not poor if it does. The poverty rate is the fraction of the population living in households that are poor.

We argue that both income and consumption measures have conceptual and measurement problems. Since the 1960s, the US Census Bureau has published an "official" poverty measure which compares cash income to a poverty threshold set in 1963. It has been heavily criticized because it uses income before taxes and transfers, excludes in-kind poverty program benefits (e.g., SNAP), and ignores costs that reduce the household's ability to purchase the minimum bundle. It is also what is called an absolute poverty measure because it uses what is called an absolute threshold, which is one held fixed in real dollars (since 1963 in this case), meaning that it does not pick up changes in how being poor is socially defined as a society develops. Use of absolute poverty thresholds also necessarily implies that, over long eras when general economic growth lifts real incomes across the income distribution, poverty rates must necessarily fall. While the magnitude of that ultimate decline is important to know, it presents an incomplete measure of socially defined well-being, at best.

The SIPM was begun by the US Census Bureau in 2009, motivated by an earlier report of the National Academy of Sciences (Citro and Michael 1995), which addresses many of the criticisms of the official measure and is widely accepted as superior to the official measure. It uses after-tax and transfer income, includes many major in-kind transfer benefits in income, and it subtracts certain costs from income as well. It uses a moving threshold based on how much it costs to purchase a minimum bundle of specifically defined necessities—food, clothing, shelter, and utilities—in the lower part of the expenditure distribution of those goods, and how that cost changes over time.

The conceptual problem with all single-period income measures is that they ignore the existence of spending out of assets and easily available borrowing, such as credit cards.[2] Conventional wisdom is that the poor, because of liquidity constraints, neither save nor borrow, so using single-period income should be accurate. We will show that, while this is true for some forms of intertemporal transfers, it is not true of all, with credit card

2. This omission is intentional and fully understood by its designers (Citro and Michael 1995, 71–72). It was argued there that current income is simply the best measure of resources and that assets are only a short-term resource.

debt being the most important. Current income does not fully represent the ability to purchase the minimum bundle if households can borrow to make such purchases, and we will show that low-income households appear to do just that. In addition to this conceptual issue, a well-known measurement issue with census income-based poverty measures is that many forms of income, particularly government transfers, are underreported in the Current Population Survey (CPS), which will tend to bias poverty rates upward.

An alternative measure which uses consumption as a measure of well-being has been proposed by Meyer and Sullivan (2012), following on work by Cutler and Katz (1991) and Slesnick (1993).[3] Many economists prefer consumption as a measure of poverty because it directly measures the flow of goods and services received by a household and therefore directly measures its economic well-being. It is also often regarded as a better measure of permanent income, which is frequently taken to be the best long-term measure of economic well-being. And, in regard to measurement, measures of consumption typically use the Consumer Expenditure Survey (CE) which is regarded by some to better measure spending than the CPS measures income.[4]

Two flaws in consumption measures make consumption a poor indicator of poverty. One is that, as agreed by all economists, a correct measure of consumption should include service flows from home, vehicles, and other durables. Yet those service flows are completely illiquid and cannot be used to purchase, with cash, food, clothing, or other components of the minimum bundle needed to satisfy basic needs. For example, almost 40 percent of low-income families are homeowners (Desilver 2021), which makes the illiquidity of housing service flows particularly important to such families. More generally, a household with a large fraction of its total consumption in the form of service flows is arguably more liquidity-constrained to buy the minimum bundle than a household with the same total consumption but which is financed entirely in the form of cash purchases.[5]

3. See Fisher and others (2009) and Fisher, Johnson, and Smeeding (2015) for additional estimates of consumption poverty.

4. The Panel Study of Income Dynamics (PSID) has now also developed enough spending measures to construct a consumption poverty measure. A comparison of its spending data to that in the CE can be found in Insolera, Simmert, and Johnson (2021).

5. In the poverty measurement literature, this is often called the problem of fungibility, meaning the degree to which some forms of income can be substituted for other forms of income. We should note, however, that the US Census Bureau includes housing in the minimum bundle, so if service flows are sufficient to satisfy the minimum housing need, that portion of the service flow is not constraining. We will discuss this below, but other durables like vehicles and household appliances are not in the minimum bundle, and hence imputing service flows to them is more potentially constraining.

The other problem with consumption measures of poverty is again related to whether intertemporal flows are possible. On the one hand, if the conventional wisdom is correct that low-income households neither save nor borrow, consumption should equal income, aside from measurement problems, and both poverty measures should produce the same poverty rate regardless of which is used because income equals consumption (Hurst 2012). But if intertemporal flows are possible—which is usually implied by the economic concept of permanent income in the first place—then consumption flows over more than one period must be included since different households may allocate their consumption differently over time. For example, a family with income just below the poverty threshold may decide to borrow on its credit card for a major purchase, raising its consumption above that threshold, while another family with exactly the same income may choose not to borrow. The first family will be counted as nonpoor and the second will be counted as poor by a single-period consumption measure, even though they have the same income and same command over resources. One family simply chooses to allocate its income to consumption in different periods than the other family.[6] Consumption in a given single period does not represent permanent income. In fact, income may be a better measure of command over resources if it is constant or fluctuating less than consumption.[7]

Our new poverty measure is intended to address both the conceptual and measurement issues with current income and poverty measures. Like most consumption poverty measures, we use data on household spending from the CE to construct our measure. However, unlike the consumption poverty concept, we consider how much a household spends to be a measure of its resources. So, for example, if a household spent $2,000 in a month, from whatever source, we simply consider that as available to spend on the minimum bundle. Almost by definition, those monies could have been spent on that bundle instead of whatever they were spent on. Using total spending as a measure of resources also differs from consumption measures

6. See Citro and Michael (1995, 210–14), who noted this issue as well.

7. We would argue that most people's intuitive definition of poverty is that it results from lack of resources, not because different families with the same resources make different choices on how to allocate their resources over time. We should also note that an old result from economic theory, called the theory of duality, states that well-being (utility) can be calculated either as a function of total resources available (using the so-called indirect utility function) or as a function of how those resources are spread across periods (e.g., discounted sum of utilities of consumption), and that the two are equivalent in their measurement of well-being. In this sense, a correct determination of available resources in each period makes an examination of consumption unnecessary and superfluous.

because the latter typically exclude spending on items that are regarded as saving and investment (e.g., cash contributions to pension plans or education and training expenses). From a resource viewpoint, those expenditures could have been spent on the minimum bundle and therefore were available to the family to have done so if they had wished and should be included in a measure of resources available.

In an important sense, our measure is closer in concept to income poverty measures because both are attempts to measure the resources available to a household. For that reason, we call our poverty measure the supplemental expenditure poverty measure (SEPM), analogous to the US Census Bureau's SIPM. But our measure of resources will exceed income if households make current purchases with credit cards that exceed their credit card debt payments or by drawing down liquid asset balances and will fall short of income if households save. If households do little of any of these activities, our expenditure poverty measure should produce poverty rates close to those of income poverty measures, apart from differing measurement error. In regard to measurement error, while many regard CE spending, in fact, to be more accurately reported, the evidence in support of that assumption is not as rigorous as one would like. There is indirect evidence that what underreporting there is in the CE is worse at the top of the income distribution (Bee, Meyer, and Sullivan 2015; Sabelhaus and others 2015; Dillman and House 2013; Attanasio and Pistaferri 2016). But there are no administrative or validated data to assess the accuracy of expenditure reporting the way there are for income reports, so most validation work, illustrated in the work just referenced, compares total expenditure reports in the CE to aggregates in the national income accounts.

Like consumption measures, using expenditures in the CE also avoids many of the constructs needed for income-based measures. We do not have to estimate taxes and tax credits, as all income-based poverty measures have to do because survey respondents cannot accurately estimate their taxes. Expenditures are, by definition, after-tax. We also do not have to impute in-kind transfers like Supplemental Nutrition Assistance Program (SNAP, food stamps) to households, as almost all income surveys have to do, because those transfers are already reflected in food expenditures reported by the household.

An important issue that has been insufficiently addressed in prior work on poverty measures but which we explicitly consider is that of liquidity. As we have already noted in our discussion of consumption poverty measures, service flows from physical assets and durables that are not part of the minimum bundle should not necessarily be considered to be available

to purchase the bundle because of their illiquidity. But a similar issue arises if current spending is treated as a resource and includes current payments on installment loans for homes (i.e., mortgage payments and interest), vehicles, and other durables purchased in the past. It would be natural to regard those as commitments from past decisions and not available for purchasing the minimum bundle in the current period. However, income poverty measures implicitly regard them as available because those payments will generally come out of current income (that is, the US Census Bureau does not deduct installment loan payments on cars, for example, from income to estimate available resources to buy the minimum bundle). Those installment loans are the result of past decisions and were therefore a matter of choice. A household could have chosen not to purchase a vehicle in the past and could have saved those monies to buy the minimum bundle in the current period. The transportation expenses in a single year are probably less than the purchase price of the vehicle, so not having purchased the car would presumably have made more funds available to buy the items in the minimum bundle net of the replacement expenses the household would have to incur. Should those past decisions and their effect on currently available resources be considered in developing a resource-based poverty measure? We will calculate poverty rates with and without some of these loan payments included in available resources as a sensitivity test.[8]

Liquidity is also important in the consideration of credit cards. Many observers see disadvantages to credit cards for low-income households because those households often do not pay off their credit card debt immediately and hence incur onerous interest rate charges which will reduce available resources in the future, and they may even default on their debt and harm their credit rating. However, low-income households subject to short-term negative consumption shocks (e.g., the car breaks down and needs a $400 repair which must be paid to be able to drive to work) and negative income shocks should find credit cards of great value to address those shocks, given their lack of cushion in other dimensions. Including credit card purchases in excess of repayments over a short period represents an important source of resources to smooth transitory shocks faced by low-income families.[9]

8. As we discuss below, the largest loan payments are those for housing. The US Census Bureau's SIPM has a special treatment for housing which, as we discuss below, we will follow. This reduces the importance of the issue to some extent.

9. Although not specifically about poor families, the Survey of Household Economics and Decisionmaking (SHED) asks how families would cover a $400 emergency expense. Of those who could not cover it with cash, credit cards are reported as the most common method (Board of Governors of the Federal Reserve System 2021). See also Fisher and Hardy (2022) for evidence on within-year volatility of consumption among the poor.

A final issue from our approach to using spending as a measure of resources is created by the implicit inclusion of spending from assets and credit card loans in our measure, since they are included in CE spending totals (but without separate identification). While we regard those as available to have been spent on the minimum bundle, it generates an inconsistent treatment between households that conduct this activity and those that do not. A household that draws down its bank balance to purchase the minimum bundle may be counted as nonpoor while another household that has the same initial balance but does not draw it down might be counted as poor. Or one household may borrow on its credit card and generate total spending in excess of the threshold and not be counted as poor, while another does not so borrow and ends up being counted as poor even though they could have borrowed (this issue is similar to that we discussed before for consumption poverty measures). To address this issue, we also calculate a resource measure that includes the potential—but unused—asset drawdown and credit card borrowing the household could have made, thereby eliminating variation in discretionary choices on how much to spend in the current period. For assets, we only include available liquid bank balances in order to restrict our measure only to easily available resources (e.g., we do not assume they could sell their car or house), and we only include credit card borrowing—not other forms of loans the household might have available—because credit cards are the easiest and most liquid form of borrowing. This liquid potential resources measure (LPRM) will represent the maximum amount of resources that are easily available to a household to purchase the minimum bundle. This maximal resource measure will count as poor households that could not buy the minimum bundle even if they pulled out every possible, easily available resource they have to do so. The LPRM will consequently count as poor those who are even more resource-deprived than those counted as poor in our main measure.

We have a number of key findings. First, we find that our main SEPM poverty rates are very close to those in the census income-based SIPM when we use the US Census Bureau's SIPM threshold. We also find that both have trended in approximately the same way (namely, downward), at least since 2010. This perhaps unexpected finding—unexpected because underreporting of CPS income and drawdown of assets and credit card borrowing should all make our CE spending totals greater than CPS income and hence our poverty rates lower—is shown to be a consequence of the precise location of the threshold combined with the differing shapes of the CPS income distribution and the CE spending distribution. Underreporting of income appears quite likely because there are many more (reportedly)

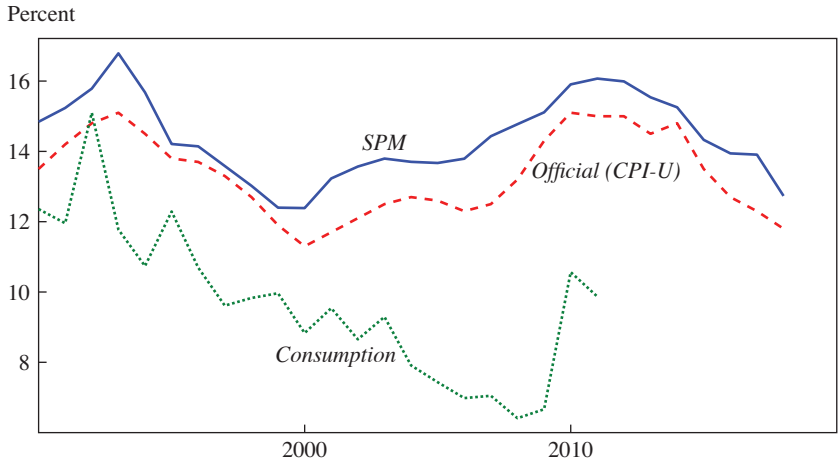
very low-income households than very low spending households. However, there is also a larger number of households with spending just below the threshold than there are households with income just below it. The two forces cancel each other out when the total number of households below the US Census Bureau threshold are counted. But thresholds just below the regular poverty threshold have more income values below the line than spending values, resulting in lower SEPM expenditure-based poverty rates than SIPM income-based rates, while the opposite occurs for slightly higher thresholds that include the nearly poor—there, SEPM poverty rates are higher than those using income by about 5 percentage points. The latter implies that there are more poor or almost poor households by expenditure than by income.

Second, we find that poverty rates for many different demographic groups are quite similar between our SEPM poverty measure and income poverty measures, with differences in the rates of less than 1 percentage point. But we find some differences between the two poverty measures that are larger than that, depending on marital status, race or ethnicity, and education level. But the largest and most notable difference occurs in poverty rates for children, where our SEPM rates are up to 2 percentage points greater than income poverty rates since 2010.

Third, we find that government transfers have a large impact in reducing expenditure poverty, by up to 5 percentage points in some years. The impact is slightly less than that implied by income poverty measures. Finally, our LPRM, consistent with conventional wisdom, shows that the liquid asset balances from bank accounts for those in the lower portion of the expenditure distribution are quite small, and their inclusion in resources has only a small effect on SEPM poverty rates. But unused and potential credit card borrowing has a greater possible impact. We find that adding these potential resources could reduce poverty rates as much as 4 percentage points. However, 9.6 percent of households, equivalent to about 31 million individuals in 2019, could still not afford to purchase the minimum bundle even after using all possible liquid resources.

The paper has three sections. The first briefly reviews previous poverty measures in the United States, with more detail than we have given in this introduction, and shows their trends reported in other work. We also describe the construction of our new measures. Section II presents our SEPM measure based solely on current expenditures and compares its level and trend to poverty rates using income measures. We also present some demographic breakdowns, including child poverty and poverty of the older population, and we show the impact of government transfer programs on

Figure 1. Official (OPM), SIPM, and Consumption Poverty Rates, 1990–2018



Sources: Fox and Burns (2021); Columbia Center for Poverty and Social Policy; Fisher, Johnson, and Smeeding (2015).
Note: Consumption rate is anchored to 2015 official threshold.

poverty rates. Section III enlarges our definition of available resources and shows its effect on poverty rates. A short summary concludes.

I. Currently Used Poverty Measures and the SEPM

We briefly review poverty rate estimates from current work on what is called the official poverty measure (OPM), the supplemental poverty measure (which we term the SIPM, with "I" for income to be analogous to the SEPM), and consumption poverty. We then present a summary of how we construct the SEPM, with details left to the online appendix.

Figure 1 shows estimates of the level and trend of poverty using three different measures after 1990. The OPM compares cash income before taxes and in-kind transfers to a threshold defined in 1963 as the amount of income needed to purchase a minimum level of food expenditure plus additional goods. It has been held constant in real CPI-U dollars since then. In addition to omitting in-kind transfers from income, it makes no adjustment for cross-area differences in the cost of living and uses a nonstandard equivalence scale to adjust for family size and composition.

The interesting aspect of the trend in OPM poverty is how little it has changed over time, despite the expectation that poverty rates should

eventually decline for any absolute poverty measure. While there are clear business cycle effects, the last value in 2018 is only slightly lower than that in 1990. In part this reflects the growth in wage inequality and the associated slow rate of growth of wages for unskilled workers. But its omission of taxes and transfers and in-kind benefits programs makes its poverty rates too high because taxes have declined for low-income families and transfers have grown over time.

The SIPM bases its threshold on a minimum bundle composed of food, clothing, shelter, and utilities and on a measure of how much is spent on those four goods in the lower part of its distribution. The threshold is updated over time as expenditures on those goods rise in that lower part, intended to represent changing social norms for where households are relative to others in the distribution of ability to purchase that bundle. This obviates the need for a price index because the threshold is defined in nominal dollars, but it implicitly picks up growth in prices of the goods in the minimum bundle. The income measure subtracts from gross money income an estimate of net taxes paid, which can be negative because of federal and state tax credits to lower-income families, and it includes estimates of in-kind transfers received by each family (SNAP plus four others noted below). The SIPM also considers working families to incur work-related expenses, which are subtracted from income, as are childcare expenses and any child support paid to a custodial parent outside the household. Somewhat more controversially, it subtracts from income a measure of medical out-of-pocket expenses, including health insurance premiums paid plus medical costs not reimbursed by insurance (Medicaid is otherwise ignored in the SIPM).[10] The SIPM also deals with homeowning by using a separate threshold for homeowners with mortgages, homeowners without mortgages, and renters, on the assumption that homeowners with mortgages need more income to purchase the rest of the minimum bundle and those without mortgages need less. It also adjusts the thresholds for a state- and metro-area level price index.

10. The total of these expenses is capped, partly because high-income families may have high medical expenses that are mostly discretionary. The latest census report describing the details of this deduction, as well as other details on how the SIPM is constructed, can be found in Fox and Burns (2021). We should note that work is currently under way to address the knotty problem of including Medicaid and health insurance in the SIPM. See Korenman, Remler, and Tyson (2019) for an important contribution on that topic. The US Office of Management and Budget (2021) also recommended that new measures adding health insurance be used to create an additional poverty index.

Given the dramatic differences in the way the SIPM and OPM are constructed from the OPM, the surprise in figure 1 is how little they differ in level and trend. The SIPM is slightly higher in level, which is not so much because of differences in the thresholds as because the subtractions from income outweigh the addition of tax credits and in-kind transfers (Fox and others 2015). The two follow similar trends over time.

Consumption poverty estimates are less standardized and differ from study to study. Those shown in figure 1 are drawn from Fisher, Johnson, and Smeeding (2015), which go through 2011.[11] The authors construct a measure of consumption which adds to nondurable spending an estimate of service flows from houses and automobiles. It also excludes expenditure items like educational expenses and pension contributions on the grounds that these constitute saving rather than consumption. The threshold used is the 2019 nominal OPM threshold, updated over time for inflation with the CPI-U-RS (after 2021 this series was renamed by BLS to R-CPI-U-RS). The consumption poverty series is lower than that of the income measures in the early years but declines at about the same rate through 2000. But after that, consumption poverty declines while income poverty rises. While consumption poverty took a large jump in 2010 (oddly, since that was the end of the Great Recession, not the beginning), its difference with the income series is dramatic.

The problem with the poverty rate estimates in figure 1, and those produced in other studies, is that they differ in too many ways to make it possible to determine why they differ. There are three basic decisions required in the construction of any poverty rate: the choice of threshold, the definition of resources, and the way the two of them are updated for inflation. The OPM rate uses a fixed real threshold established in 1963 and a narrow definition of resources and updates with the CPI-U. The SIPM uses a threshold that is adjusted in real terms over time (generally upward) and a more comprehensive definition of income and implicitly uses a price index for food, clothing, housing, and utilities. The consumption measure uses a constant real threshold similar to the OPM and imputed service flows for durables and employs the CPI-U-RS for price updating. Because none of these studies analyze which of these three building blocks is responsible for the differences in level and trend, the reason for their differences cannot be determined. One goal of our study is to compare our SEPM poverty series

11. Meyer and Sullivan (2019) have the latest consumption poverty series using their methodology, but they use a very different price index than other studies, making it noncomparable to the other series in figure 1.

to that of the SIPM on a comparable basis so that we can determine exactly what difference is made by using spending instead of income alone, and at least address this issue with two of the poverty measures.[12]

Turning to the construction of SEPM, our basic SEPM poverty measure uses consumer expenditure from the CE as the building block of available resources.[13] We do not exclude any items that might be regarded as investment or saving because those could have been used, instead, to buy the minimum bundle and hence should be included in resources. We also include all down payments on durables in our expenditure measure, because the household could have chosen not to purchase the durable in question and could have applied that expenditure toward the minimum bundle instead. For installment loans, the CE only collects data on such payments for housing and cars and not on those for any other durables.[14] We include outlays for both in our measure of expenditures on the grounds that those are cash payments and are therefore liquid. We recognize that their inclusion could be objected to on liquidity grounds but, unlike service flows, they represent actual cash outlays that could in principle have been redirected toward the purchase of the minimum bundle if the debt had not been incurred in the first place.[15] However, we include installment loan payments on houses (i.e., mortgage payments) in our expenditure measure for a second and independent reason, which is that this is required to be comparable with the SIPM treatment of housing. The SIPM recognizes the importance of housing to low-income families and that treating homeowners the same whether or not they have a mortgage, and the same as renters, misrepresents differences in implicit income and hence ability to purchase the minimum bundle (and housing is in the minimum bundle). On the grounds that estimating service flows is too difficult as a practical matter, given existing data and methods, the SIPM instead adjusts the threshold upward for homeowners who have mortgages and downward for

12. Constructing a new consumption poverty measure is beyond the scope of our paper and is left for future work.

13. We note that the CE uses the word *outlays* for our measure. We use the word *expenditures*, which is more commonly used outside the CE.

14. For other durables, such as refrigerators, dishwashers, and washing machines and dryers, for example, the CE just includes in spending the purchase price at the time of purchase and ignores whether they are purchased on credit. We note that Bruce Meyer's comment in the general discussion—that we exclude consumption on housing and transportation—is incorrect because we include all cash expenditures on those goods. We do not include illiquid implicit service flows.

15. Online appendix 1 describes many of the details involved in implementing these decisions. The CE only includes purchase price for some durables, even if financed by a loan, which we can do nothing about.

homeowners who do not have mortgages on the grounds that the former group needs more income to be able to purchase the nonhousing items in the minimum bundle and the latter need less. Thresholds for renters are adjusted based on average rents paid by lower-income renters, consistent with the notion of a socially defined threshold for low-income households. With this adjustment of the threshold, mortgage payments must be included in any resource measure, including our expenditure construct.[16] However, we conduct a sensitivity test to the inclusion of vehicle loan payments in spending, reported in the online appendix.

For credit cards, we have emphasized that purchases made with credit cards are implicitly included in the CE expenditure measure, although the respondents are not asked how many purchases are actually made with cards and hence those purchases cannot be separated from purchases made from other resources. In addition, the CE does not ask households about their interest and fees on credit cards in every interview nor does it ask the amount by which households pay down their credit card balances. However, fortunately, the CE excludes credit card interest, fees, and debt payments from its expenditure measure, so they are not counted in our expenditure totals. Since purchases made with credit cards are implicitly included in our spending measure, this means that any household which pays off its credit card balances every period will have no greater calculated available resources than a household which makes no credit card purchases; the net will be zero in either case. But households that make purchases in excess of their interest, fees, and debt payments will be implicitly regarded as having additional resources, and the opposite will be the case for households whose new charges are less than their interest, fees, and debt payments.[17] The annual time frame for our SEPM and most other poverty measures makes this an internally consistent approach.[18]

16. See Fox and Burns (2021) for details. The threshold is adjusted only for the housing cost portion of the minimum bundle. Implicitly, this treatment subtracts from income any housing expenditures deemed necessary to purchase the housing portion of the minimum bundle, leaving remaining income to purchase the rest of the bundle (and other things, including more housing). We note that this treatment of housing therefore partly reduces the problem of putting illiquid housing service flows into income. A recent commission in the United Kingdom has also recommended that mortgages be subtracted from income for poverty measurement (Social Metrics Commission 2020).

17. Our original conference paper proposed counting both credit card spending and repayment as expenses, as is noted in the general discussion of our paper. This revision eliminates that double counting.

18. If the net adds to zero across the population (i.e., the sum of new charges in excess of interest, fees, and debt payments equals the sum of new charges less those items), as will be the case if some households are net creditors and others are net debtors in different periods,

Because we want to make the SIPM our main poverty measure of comparison and to have our measure as comparable to it as possible save for the use of expenditures in place of income, we adopt all other methods used by the census in constructing that measure. We use the same thresholds as the SIPM, the same differentiation of those thresholds by homeowner and mortgage status (as already noted), the same type of geographic cost-of-living adjustments, and the same family size equivalency scale used in threshold construction.[19] We also add to our expenditure total estimated amounts of the four in-kind transfers other than SNAP which the SIPM adds to income and which are not recorded as expenditures in the CE: implicit rent subsidies to those in government-subsidized housing who pay below-market rents, lunch subsidies received by schoolchildren, transfers in federal nutrition programs for pregnant women and mothers of young children, and energy assistance. We recognize that liquidity issues can be raised with these estimates as well and hope that they will be small enough in magnitude as not to constrain the family in its ability to purchase the minimum bundle. Finally, like the SIPM does for income, we also deduct from our expenditures work-related and childcare expenses, child support paid, and capped out-of-pocket medical expenses, though all necessarily must be computed with CE data instead of the CPS. These adjustments are an important feature of both the SEPM and SIPM poverty measures.[20]

One issue with the CE worth noting is that the CE data are collected in quarterly interviews, not annual interviews like the CPS Annual Social and Economic Supplement (ASEC). In the construction of annual totals, the Bureau of Labor Statistics (BLS) treats each quarter as an independent observation and then averages them with weights to arrive at calendar year estimates. This approach contrasts with some authors who use only a subsample—for example, Bavier (2014), who uses only the Q2 interview—or authors who use only households that complete all interviews—for

---

then the impact of credit cards on poverty rate estimation depends only on the distribution of the two types of households in the region of the poverty threshold where households are moved either above it or below it by the inclusion of their net values in resources. We thank Henry Aaron for making this point.

19. We thank Caroline Hoxby for noting that low-income families are members of networks that share resources and consumption, including family members outside the unit, neighbors and friends, absent fathers, and others. These networks could also be the source of some of the additional spending over income found in low-income household data. Spending which arises from outside the family unit will be included in our measure of spending but would be excluded by an income-based measure.

20. See the online appendix for details on the implementation of these procedures with the CE data.

example, Fisher, Johnson, and Smeeding (2015). If a sample of consumer units present in all four quarters is required, significant sample loss occurs from attrition, for about 45 percent of the sample leaves the survey. Further, attrition is non-ignorable because those remaining in the sample are more educated, more likely to be homeowners, more White, and more elderly, and thus less likely to be poor. Given the difficulties in correcting for attrition, we follow BLS in constructing annual expenditures from quarterly amounts, but this may have some effect on calculated poverty rates because quarterly expenditure may fluctuate more than annual expenditure. In this case, our SEPM poverty rates may be higher than those from an annual measure like the SIPM to some extent. We leave this issue for future work.

Finally, we will construct a "maximal" estimate of resources by expanding the definition of total available resources to include liquid assets and potential liquid borrowing in our calculations. We calculate our measure of liquid potential resources as

$$LPR = Current\ Expenditures + Additional\ Available\ Liquid\ Assets + Additional\ Available\ Liquid\ Borrowing$$

We use data on current savings and checking bank balances at the end of the year recorded in the CE to calculate additional available liquid assets.[21] Calculating additional potential credit card borrowing is more difficult both for data and conceptual reasons. For those with credit cards, calculating additional borrowing potential requires knowing current balances plus credit limits on those cards, and the CE asks credit card balance information but not limits. We need to use other data for limits and impute those to CE households. Traditional credit card rating agencies have data on limits but not income, which is needed to identify low-income households. We use the Survey of Consumer Finances (SCF)—a representative survey of US households focusing on financial information—which has data on income, credit card usage, and credit card limits. We impute credit card limits from the SCF to the CE using methods (which are based on income and age strata) described in the online appendix, and we calculate unused credit as the imputed limit minus the balance reported in the CE. We impute to those who report zero CE credit card balances some fraction to have a card (the CE does not ask if households have a card), again from the SCF, and credit limits to those households, for whom unused credit equals the limit.[22] Because of the large number of imputations necessary to construct unused

21. The CE only collects these data in the last quarterly interview, so we must restrict our sample to non-attriting households for this calculation.

22. See the online appendix for details.

credit, given the available data, our calculations should only be considered as suggestive.

We recognize that if the household were to draw its full potential in the current period, it would reduce its potential resources in future periods. It cannot draw those resources down period after period. But this is a consequence of the annual time frame used in most resource measures (including income poverty measures that include saving in resources). Annual time frame poverty measures ask only whether resources in a current year are large enough to buy the minimum bundle in that year, not whether resources over multiple years are large enough to buy the minimum bundle repeatedly. But an interesting extension of the standard annual measure would be to estimate the current value of assets and borrowing taking into account their impact in constraining future ability to purchase the minimum bundle or, phrased differently, how available resources vary as the time frame lengthens.

II. Results

II.A. Levels in 2017–2019

Before comparing trends in our SEPM to that of the SIPM, we present levels of the two measures averaged over our last three years of data, 2017–2019, to illustrate the building blocks for each and the nature of their construction (averaged over three years to smooth out short-term fluctuations in the measures). We also present a first major finding on the relationship between our expenditure poverty measure and income measures in this initial exercise.

Table 1 shows the building blocks for our SEPM using the CE and the SIPM using the CPS for 2017–2019. The first rows present statistics on the distributions of gross CE expenditure and gross adjusted CPS income.[23] In the whole population, CE mean and median expenditures are much lower than for income in the CPS, but this deserves little attention because it is the lower tails of each that are relevant to poverty measurement. However, an important result in the table is that the income distribution in the CPS has a much longer left-hand tail than the expenditure distribution in the CE, and the difference gets larger the lower in the distribution one goes. The best explanation for this is simple underreporting of income in

23. Adjusted CPS income is CPS income after tax and with the most important in-kind transfer—SNAP—added. This is a closer concept to CE spending than before-tax cash income and should improve comparability relative to using before-tax CPS cash income alone.

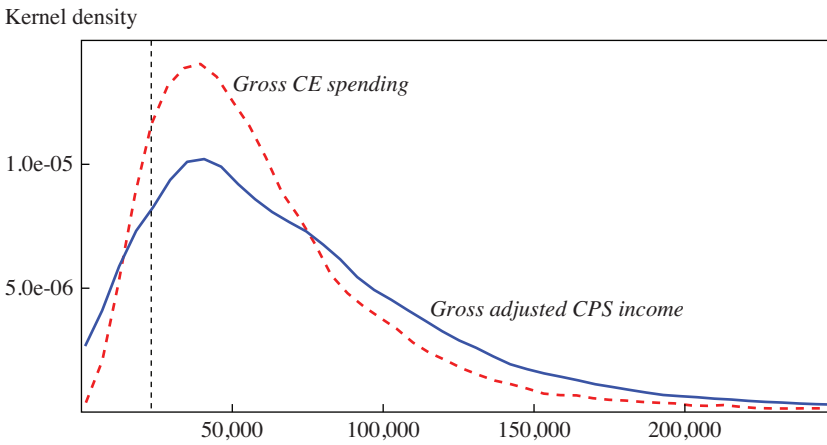
Table 1. CE SEPM and CPS SIPM Poverty Rates and Components, 2017–2019

	CE		*CPS*	
	Statistic	*SE*	*Statistic*	*SE*
Gross expenditure or gross adjusted income				
Mean	62,957	192	78,268	160
Median	51,628		61,672	
1st percentile	9,436		0	
3rd percentile	13,654		6,947	
5th percentile	16,542		11,245	
10th percentile	21,662		18,629	
20th percentile	29,596		30,489	
Net expenditure or net adjusted income				
Mean	56,251	185	70,711	156
Median	44,605		53,422	
1st percentile	8,047		0	
3rd percentile	11,634		5,204	
5th percentile	14,013		9,191	
10th percentile	18,472		15,546	
20th percentile	25,389		25,551	
Poverty rates				
Gross SEPM or SIPM	0.089		0.096	
Net SEPM or SIPM	0.133		0.130	
Means adjustments and in-kind in bottom quintile of the distribution				
Adjustments				
Medical out-of-pocket spending	2,911	23	2,632	12
Work expenses and childcare	798	10	986	5
Child support	20	1	42	2
Total adjustments	3,729	24	3,660	13
In-kind transfers				
School lunch subsidy	198	3	241	2
Energy assistance	29	1	40	1
WIC	55	2	51	1
Housing subsidy	897	20	786	11
Total in-kind	1,179	22	1,118	12
Demographics				
Family size	2.453	0.012	2.267	0.006
Children	0.741	0.009	0.695	0.005
Adults	1.713	0.006	1.572	0.003
Presence of elderly	0.304		0.286	
Own with a mortgage	0.151		0.155	
Own no mortgage	0.307		0.305	
Renters	0.542		0.539	
Sample size	62,867		205,618	

Source: Authors' calculations.

Note: Values are expressed in 2014 dollars. Gross expenditure is total household spending on all items in the year. Gross adjusted income is total income in the year after tax and with SNAP benefits added. Net adjusted income includes four in-kind transfers and excludes three types of capped adjustments. Poverty rates weighted by person, household weighted by consumer unit weight. See online appendix for details.

Figure 2. Distribution of Gross CE Spending and Gross Adjusted CPS Income, 2017–2019



Source: Authors' calculations.

Note: Gross CE spending is total household spending on all items in the year. Gross adjusted CPS income is total income in the year after tax and with SNAP benefits added. Vertical dashed line denotes average threshold.

the CPS, but whatever the cause, it implies that poverty rates may differ simply because of this difference, as we now illustrate.[24]

Figure 2 shows the two distributions graphically but in dollar terms and not percentile terms. The vertical dashed line shows the average SIPM threshold (approximately $26,000 in 2019) so that poverty rates can be viewed as the fraction of the distribution to the left of that line. The most important difference, as suggested by table 1, is that expenditures are much more concentrated in a mass just above the threshold, unlike the more dispersed income distribution. Because the density curves cross and hence neither distribution first-order stochastically dominates the other, the relative poverty rates of the SEPM and SIPM will depend on where the threshold is located. In figure 2, it is not visually apparent whether expenditures or income have a greater fraction to the left of the line. But table 1, showing gross SEPM and SIPM poverty rates, shows that the percentage of reported income observations below the threshold, 9.6 percent, is slightly higher but

24. We thank our discussants, Kathryn Edin and Luke Shaefer, for noting that the gap may not be entirely a result of literal underreporting of income and income transfers but rather partly reflecting the adoption of (costly) survival strategies by low-income families to find ways to obtain more consumption in light of incomes too low to survive.

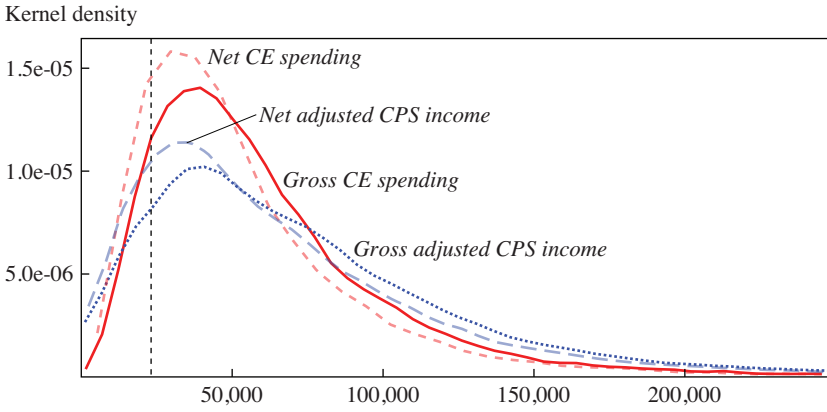
very close to the fraction of expenditures below the threshold, 8.9 percent. Thus, the differences in the distributions of income and expenditure below the poverty line almost cancel out.

As we noted above, the US Census Bureau SIPM adds certain in-kind transfers to income and subtracts certain adjustments representing costs before calculating ability to purchase the minimum bundle. What we term the net poverty rate is that based on net expenditure and net income after these additions and subtractions. Table 1 shows the distributions of net expenditure and net income, in parallel to those for the gross distributions. Not surprisingly, we continue to find a longer left tail of net income than net expenditures, which should be the case if the in-kind transfers and deducted adjustments are roughly the same in the two data sets. The means of those in-kind transfers and deducted adjustments are shown in the lower half of the table and demonstrate that their means are not much different in the CE and CPS.

However, the relationship between the two poverty rates changes slightly when going to net expenditures and income. Both the SEPM and SIPM net poverty rates are higher than their gross counterparts because the deductions for cost factors are larger than the additions from in-kind values. However, the SEPM rises more than the SIPM (4.4 percentage points compared to 3.4 percentage points), resulting in an almost identical net poverty rate for the two—13.3 percent for the SEPM and 13.0 percent for the SIPM. The major reason for the change is illustrated in figure 3, which adds the distributions of net expenditure and net income to those for their gross counterparts which were shown in figure 2. Both distributions are shifted to the left, but because of the greater mass of the gross expenditure distribution just above the threshold, more household expenditures are moved below the threshold than are household incomes, when netting out the cost factors.

The important lesson for poverty measurement is that the relationship between income and expenditure poverty rates depends critically on where the threshold is fixed. Since all observers agree that the choice of threshold is socially determined and has arbitrary elements, most observers think that poverty rates at different thresholds should be calculated. Figure 4 shows one such calculation, illustrating the importance of the threshold by showing net SEPM and SIPM poverty rates for what are designated, in the literature, "deep poverty" and "near poverty." The first is calculated as the fraction of the population which has income or expenditure less than 50 percent of the threshold, and the latter is calculated as the fraction of the population which has income or expenditure less than 150 percent of the threshold (approximately $13,000 and $39,000, respectively in 2019).

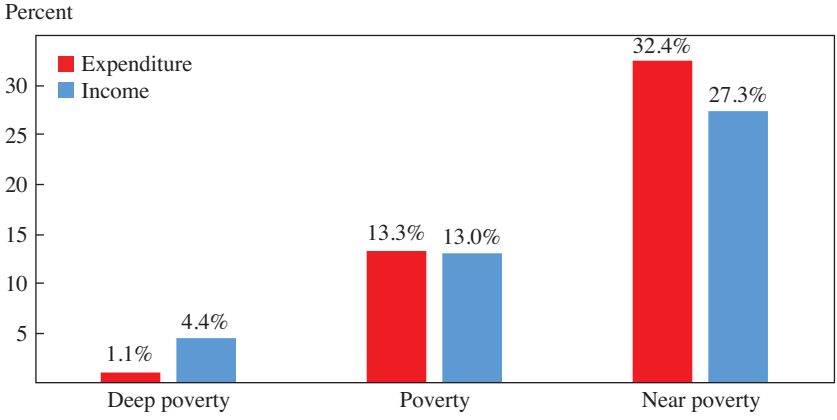
Figure 3. Gross and Net CE Spending and Adjusted CPS Income, 2017–2019



Source: Authors' calculations.

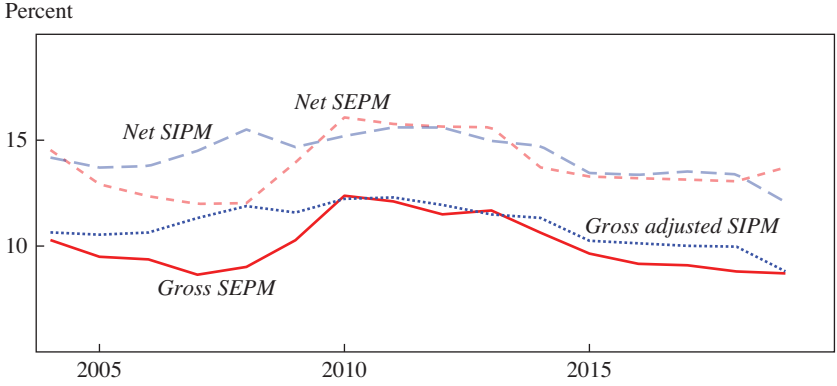
Note: Gross CE spending is total household spending on all items in the year. Gross adjusted CPS income is total income in the year after tax and with SNAP benefits added. Net measures include four in-kind transfers and exclude three types of capped adjustments (work-related and childcare costs, child support paid, and out-of-pocket medical expense). Vertical dashed line denotes average threshold.

Figure 4. SEPM and SIPM Net Poverty Rates by Threshold Location



Source: Authors' calculations.

Figure 5. SEPM and SIPM Poverty Rates, Gross and Net, 2004–2019



Source: Authors' calculations.

Note: Gross SEPM poverty rate is based on total household spending on all items in the year. Gross adjusted SIPM poverty rate is based on total income in the year after tax and with SNAP benefits added. Net poverty rates are based on total spending and income after tax and with SNAP benefits that include three in-kind transfers and exclude three types of capped adjustments (work-related and childcare costs, child support paid, and out-of-pocket medical expense).

The figure reveals that SEPM net poverty rates are lower than those for the SIPM when looking at deep poverty, but higher than those for the SIPM when looking at near poverty. There is more SIPM deep poverty than SEPM deep poverty (4.4 percent versus 1.1 percent) but more SEPM near poverty than SIPM near poverty (32.4 percent versus 27.3 percent, a 5 percentage point difference of about 16 million individuals). There are very few households with extremely low expenditures but a large fraction of households with expenditures that are still fairly low. There may be more households that are very poor by income standards, but there are also many households that are almost poor by expenditure standards. The latter group should not be considered particularly well-off in terms of economic resources.

II.B. Trends, 2004–2019

Trends in gross and net SIPM and SEPM poverty rates from 2004 to 2019 are shown in figure 5. We show both net and gross rates since there are some differences between them, as there were in 2017–2019. The gross SEPM poverty rate was approximately 11 percent in 2004, fell to about 8 percent in 2007, then rose through 2010 to about 12 percent (no doubt because of the Great Recession). It then began a gradual decline to a 2019 value of 8.7 percent (the decline coinciding with a general economic growth period in the country). The gross SIPM poverty rate shows higher values

in the 2004–2007 period, a somewhat sharper rise from 2007 to 2011, and then a sharper fall through 2019, ending at its final value of 8.8 percent, almost identical to that for the SEPM.[25]

The difference in the measures from 2004 to 2010 has been noted before although not using quite the same income and expenditure poverty rates we calculate (Bavier 2014; Wimer 2014). The difference has not been resolved, but the SEPM exhibits a pattern more consistent with the business cycle in this period—strong economic growth from 2004 to 2007 followed by the Great Recession from 2007 to 2010, which is consistent with falling then rising poverty rates—than the SIPM. However, from 2010 to 2019, both the gross SEPM and SIPM follow approximately the same downward trend on average. The economic growth over this period is the likely cause of both, together with expanded social safety net transfers. When moving to the net poverty rates, both the SEPM and SIPM shift upward, as already discussed, but the shift upward results in a similar pattern of time trends of each over the entire 2004–2019 period. Both have continued to decline since the Great Recession, as was the case for the gross measures.[26]

Given the importance of the location of the threshold, we show trends in deep poverty and near poverty in figure 6. The greater rates for SIPM than for SEPM for deep poverty have been present since 2004, and both show very flat trends with very little reduction in the rates. The lack of improvement in deep poverty rates is a result of a combination of declining labor market earnings at the bottom of the distribution and a decline in transfers going to the worst-off families. Near poverty SEPM rates were not higher than those for SIPM over the whole period but have been for most of it, and both show approximately the same declines since about 2010.

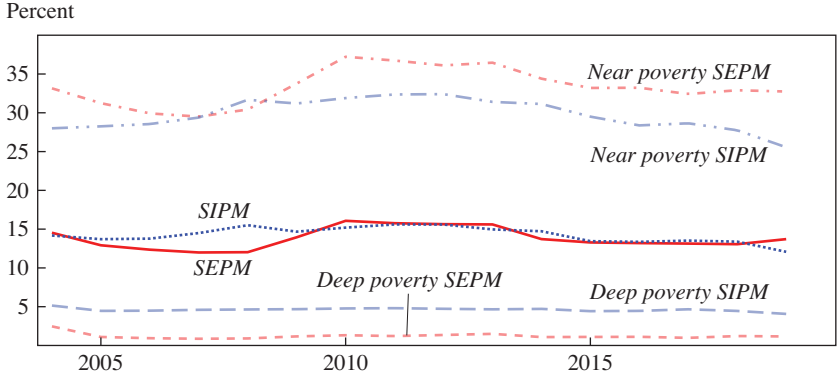
II.C. Comparison of Demographic Patterns

Table 2 shows SEPM and SIPM gross and net poverty rates in 2017–2019 for different demographic groups to determine whether the two measures yield different rates. Different poverty rates can arise for the same reason

25. Our SIPM poverty rates are calculated from the public-use CPS historical files produced by the Columbia Center for Poverty and Social Policy (CPSP). Those rates differ slightly from those produced by the US Census Bureau since 2009. Our rates also differ slightly from those produced by the CPSP because we modify their procedures for medical and work expense imputations, geographic adjustments, and household weights, as described in the online appendix.

26. The uptick in the net SEPM measure in 2019 is largely a result of stagnant net expenditure spending in CE from 2018 to 2019 but a rise in the SIPM threshold, resulting in higher poverty. The fall in net SIPM in 2019 reflects a significant rise in net incomes in CPS ASEC from 2018 to 2019.

Figure 6. Near and Deep Net Poverty Rates, 2004–2019



Source: Authors' calculations.

Note: SEPM based on adjusted expenditure below the SPM threshold. SIPM based on adjusted income below the SPM threshold. An individual is in deep poverty if their adjusted resources are less than 50 percent of the poverty threshold. An individual is in near poverty if their adjusted resources are less than 150 percent of the poverty threshold. Resources include SNAP benefits and three in-kind transfers and exclude three types of capped adjustments (work-related and childcare costs, child support paid, and out-of-pocket medical expense).

Table 2. Poverty Status by Demographic Groups, 2017–2019

	SEPM Gross	*SEPM Net*	*SIPM Gross*	*SIPM Net*
Home				
Owner w/mortgage	0.027	0.052	0.035	0.057
Owner w/o mortgage	0.082	0.154	0.079	0.119
Renter	0.171	0.219	0.191	0.237
Family type				
Unmarried	0.141	0.187	0.167	0.215
Married	0.059	0.102	0.053	0.079
Poverty status by age				
Elderly poverty rate	0.069	0.150	0.096	0.158
Child poverty rate	0.118	0.160	0.111	0.135
Race and ethnicity				
White	0.053	0.092	0.062	0.088
Black	0.153	0.188	0.172	0.213
Hispanic	0.167	0.236	0.160	0.216
Other	0.097	0.138	0.109	0.147
Education				
Less than high school	0.259	0.333	0.254	0.325
High school	0.106	0.161	0.115	0.156
College degree, including associates	0.029	0.055	0.046	0.065
Poverty rate	0.089	0.133	0.096	0.130
Sample size	62,867		205,618	

Source: Authors' calculations.

Note: Demographic characteristics refer to the household reference person. The sample is weighted by person weights. See online appendix for details.

already noted, which is simply that the distributions of the two are different below the SIPM threshold and some groups may have more expenditures just below the poverty threshold than others. Differences in net poverty rates can also differ, in principle, if the values of the in-kind transfer additions or the deduction subtractions are different for some unknown reason in the CPS and CE, although we have not found those to be dramatically different in the two data sets. The last few rows in table 1 show that there are some differences in a few demographic variables in the CPS and CE.

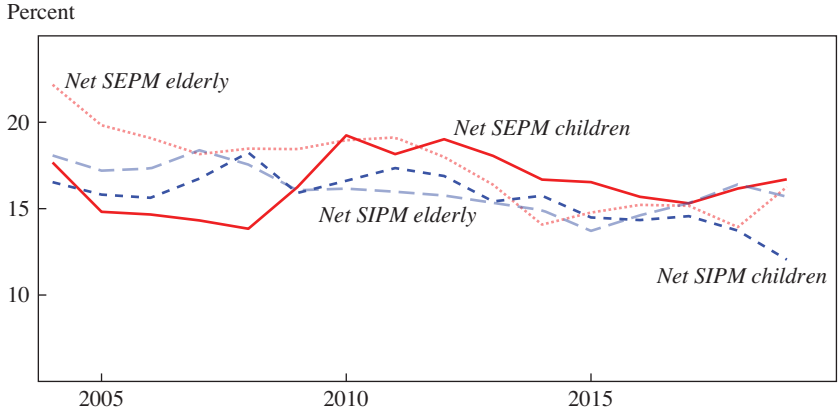
Most of the differences in SEPM and SIPM poverty rates in table 2 are not large by demographic characteristic, often less than 1 percentage point and varying in which poverty rate is the higher. But there are a few differences that are more than 2 percentage points. Owners without a mortgage have a net poverty rate over 3 percentage points greater for the SEPM than for the SIPM, while renters have a lower SEPM poverty rate. SEPM poverty rates are lower for unmarried households than for the SIPM, but the opposite is the case for married households.[27] Elderly persons have about the same gross SIPM and SEPM poverty rates. For children, the net SEPM is higher than the net SIPM. Black households, but also households with heads who have a college degree, have lower SEPM poverty rates than those for income.[28]

However, one major difference in the rates between the two measures is for children. Child poverty rates have always been calculated to be higher than those for adults by all poverty measures because more children tend to live in poor families. Figure 7 shows trends from 2004 to 2019 in net SEPM and SIPM poverty rates for children and the elderly. While those for the elderly are, on average, quite close to one another, consistent with the 2017–2019 average result in table 2, the SEPM child poverty rates are much higher than SIPM poverty rates since 2010. At their peak in the period 2010–2013, SEPM child poverty rates were almost 19 percent, about two percentage points higher than rates based on income. This reflects the greater concentration of expenditures of households with children just below the poverty threshold. However, child poverty rates have also declined over time.

27. It should be emphasized that family size is taken into account in the determination of the thresholds, so they differ, for example, for single individuals and married individuals.

28. Online appendix table A1 shows differences in various characteristics for the SEPM and SIPM poor. As expected, the SEPM poor have higher expenditures than the SIPM poor have income. There are a few demographic differences as well. For example, the SEPM poor have larger family sizes, and the household reference person has less education.

Figure 7. Net SEPM and SIPM Poverty Rates, Children and Elderly, 2004–2019



Source: Authors' calculations.

Note: Net poverty rates are based on total spending and income after tax and with SNAP benefits that include three in-kind transfers and exclude three types of capped adjustments (work-related and childcare costs, child support paid, and out-of-pocket medical expense).

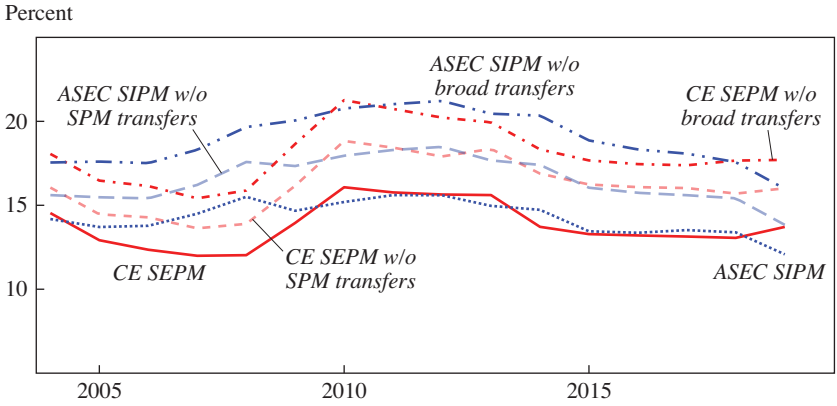
Online appendix table A1 shows a comparison of the demographic characteristics of SIPM poor and SEPM poor to determine whether they identify the same or different types of households as poor. The SEPM poor and SIPM poor are not, in fact, very different by the majority of measures shown in the table. The few larger differences include a greater fraction of lower-educated household heads among the SEPM poor as well as larger family sizes. There are also differences in how many homeowners have and do not have a mortgage. But these are the exceptions rather than the rule.[29]

II.D. Impact of Government Transfers

Assessing the impact of government transfers with our expenditure measure requires assumptions not needed for assessing that impact with income measures. With income used as a measure of transfers, transfers represent a simple addition to income and hence a straightforward calculation of their impact on poverty rates can be conducted (ignoring behavioral responses). But for expenditure measures, an assumption is needed on how an increase

29. As noted by our discussants, Kathryn Edin and Luke Shaefer, it would be useful to know how the two measures classify as poor families with particular material hardships, such as food insecurity or defaulting on rents, mortgages, or utility bills. Unfortunately, the CE does not have information on measures of hardship.

Figure 8. Impact of Selected Transfers on Poverty, All Persons, 2004–2019



Source: Authors' calculations.

Note: SIPM measures use TAXSIM, geographic cost-of-living adjustment, and 15 group imputation. SPM transfers are SNAP, WIC, housing assistance, energy assistance, and school lunch. Broad transfers are SPM transfers plus the EITC, cash welfare, and SSI.

in income is spread out across expenditures in different periods, as well as an assumption of whether drawdowns from liquid assets or credit card borrowing are affected. The simplest assumption is to assume that neither of the latter are affected by transfers and that all transfers result in increased expenditure in the current period.

With that assumption, figure 8 shows the impact on net SEPM and SIPM poverty rates when transfers are removed. In the first case, we consider the impact on poverty rates of removing the in-kind transfers alone—SNAP and the four others discussed previously. In the second step, we consider the impact on poverty of removing cash transfers, which is primarily the Earned Income Tax Credit (EITC) but also cash welfare and Supplemental Security Income (SSI) (both smaller in magnitude than the EITC). Focusing on the period since 2010, we find that the removal of in-kind transfers raises both SEPM and SIPM poverty rates by approximately 3 percentage points, with no large difference, on average, between the measures. But we also find that removing cash transfers increases the SEPM less than the SIPM. Online appendix table A2 shows that the removal of SSI, cash welfare, and other in-kind transfers has a greater impact on raising SIPM poverty than SEPM poverty. In any case, however, figure 8 shows that taxes and transfers to disadvantaged families in the United States make a major dent in poverty rates even with our SEPM, up to 4 or 5 percentage points.

II.E. Liquid Potential Resources

As described previously, we estimate an LPRM by adding available liquid bank balances and an estimate of available but unused credit card borrowing resources to obtain a maximal measure of resources and to estimate how many individuals would remain in poverty even after using all available liquid assets and credit. Table 3 shows the mean and median bank balances (liquid assets) at the final interview for households in the bottom quartile of the CE current expenditure distribution for 2017–2019, shown separately by the three housing statuses employed by the US Census Bureau in its threshold calculations, and also broken out by whether the household head is or is not age 65 or older. Median bank balances are zero for those with heads under age 65 and small for those over age 65 but a substantial fraction, sometimes over 50 percent, have a positive bank balance. Mean assets are small for those under age 65 but quite large for some of those over age 65, but this reflects a large upper tail of the distribution.

Figure 9 shows the impact of adding these bank balances to available resources on the SEPM net poverty rate, in level and trend. The rate declines by about 1.5 percentage points on average. There is very little change in the impact over time from 2004 to 2019, reflecting little change in the amount and distribution of bank balances relative to total household spending.[30]

Table 3 also shows estimates of unused credit calculated with two different imputation methods from the SCF, with little difference between them (see the online appendix). For the main method, median unused credit is again zero or small for households with heads under age 65 but often sizable for households with elderly heads. Mean unused credit is again much higher than median values, reflecting right-skewed distributions, and is generally modest for the non-elderly but greater for the elderly. Figure 9 shows the impact of adding unused credit to resources in addition to bank balances (using the main method), showing that poverty rates are reduced by about 3 percentage points from this addition, a nontrivial reduction. Given the crudeness of the estimates, this is only a rough estimate, but it does establish the potential importance of the issue. Interestingly, the impact varies little over time, implicitly meaning that credit card non-utilization (in dollar terms) has not changed very much for lower-income families.[31] Nevertheless, the poverty rate, even if all bank balances were used and all

30. Results omitting the elderly (available upon request) are very close to those in figure 9.

31. The approximate constancy of unused credit is a result of offsetting fluctuations from year to year in the proportion with cards, the proportion of those with cards but zero balances, and the utilization rate.

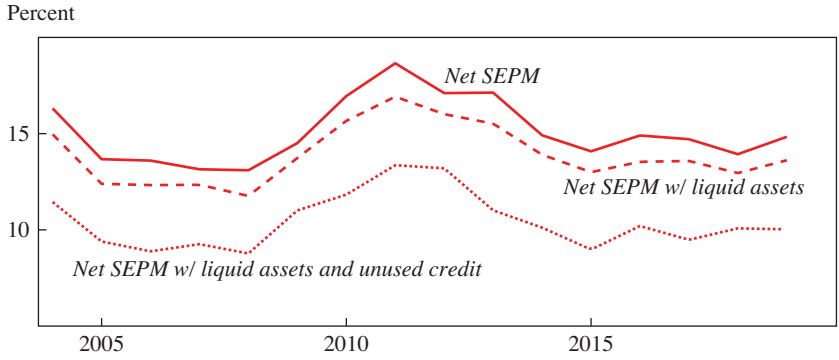
Table 3. Liquid Assets and Unused Credit for Bottom Quartile of Households, 2017–2019

	Under age 65	*Age 65 or older*	*Total*
Owners w/mortgage			
Median			
Liquid assets	0	47	0
Unused credit	506	4,625	1,222
Unused credit, alternative imputation	0	3,039	565
Mean			
Liquid assets	1,011	4,346	1,966
Unused credit	2,862	6,264	3,836
Unused credit, alternative imputation	2,277	4,002	2,771
Positive liquid asset balance (%)	44.5	54.0	47.2
Positive credit balance (%)	22.3	29.8	24.4
Owners w/o mortgage			
Median			
Liquid assets	0	370	1
Unused credit	352	4,625	1,665
Unused credit, alternative imputation	675	4,625	2,412
Mean			
Liquid assets	3,811	20,038.0	11,852
Unused credit	3,062	5,369	4,205
Unused credit, alternative imputation	3,093	5,065	4,070
Positive liquid asset balance (%)	43.2	58.3	50.7
Positive credit balance (%)	17.4	19.2	18.3
Renters			
Median			
Liquid assets	0	1	0
Unused credit	0	476	0
Unused credit, alternative imputation	0	541	0
Mean			
Liquid assets	739	2,371	970
Unused credit	1,518	3,550	1,805
Unused credit, alternative imputation	1,479	3,170	1,718
Positive liquid asset balance (%)	37.2	50.5	39.1
Positive credit balance (%)	12.3	13.6	12.5
Frequency	4,365	1,893	5,096

Source: Authors' calculations.

Note: Sample is composed of the bottom quartile of the gross expenditure distribution. Having a credit card is imputed based on income and age groups. Unused credit is the difference between an individual's imputed limit and their balance. Credit limits are imputed based on income, age, and credit balance. The alternative imputation of credit limits uses only income and age groups. The sample is weighted by person weights. See online appendix for details.

Figure 9. Net SEPM Poverty Rates with and without Liquid Assets and Unused Credit, 2004–2019



Source: Authors' computations.

Note: SEPM based on adjusted expenditure below the SPM threshold. Resources include SNAP benefits and three in-kind transfers and exclude three types of capped adjustments (work-related and childcare costs, child support paid, and out-of-pocket medical expense). SEPM with geographic cost-of-living adjustment is used. Liquid assets include checking and savings account balances. Unused credit is imputed from the SCF based on income, age, and credit utilization rate.

available credit were utilized, is still 9.6 percent in 2019, leaving almost 31 million individuals still in poverty and without the resources to meet basic needs. This constitutes a particularly resource-deprived group of poor families.

III. Summary and Conclusions

This paper has proposed a new poverty measure that we argue has advantages over income poverty and consumption poverty measures. Our measure is based on observed, realized spending as a measure of the resources available to a household, either alone or supplemented with access to resources from bank balances and credit cards. We argue that it has advantages relative to income measures because it includes in resources spending from credit cards and spending out of liquid bank balances, and it is superior to consumption measures because it does not count illiquid service flows from housing and vehicles as resources and better accounts for households that allocate their consumption differently across years. Empirically, it is preferable to income if CE expenditures are measured more accurately than income in surveys like the CPS. Our measure also has several practical advantages over income poverty measures because it does not require

estimation of taxes, adjustments for underreporting of transfers, or the imputation of some in-kind transfers.

We implement our SEPM on the CE data from 2004 to 2019. We find that SEPM poverty rates—based just on total household expenditures in a period—were nearly the same in 2017–2019 as those estimated with income data from the CPS. However, expenditure poverty rates depend critically on exactly where the poverty line is drawn because there is a large mass of households with expenditures only just above the most widely accepted threshold used by the US Census Bureau. Moving the poverty line up slightly to capture those households who are almost poor but not quite poor makes SEPM poverty rates 5 percentage points (about 16 million individuals) higher than those using income. Overall, we find that there are many more low-expenditure households in the United States than low-income households, in percentage terms.

We also assess the ability of households to escape poverty by drawing on available liquid bank balances and by using available, but unused, credit debt to finance purchases of basic goods. Many low-income households already do that, but some do not use all the potential borrowing they could. We find that bank balances are quite small and, when counted toward ability to escape poverty, make only a small difference in reducing poverty rates. But we find that available credit card borrowing could potentially lower poverty rates further by up to 3 percentage points. However, the arguably most important finding is that even if households were to draw down their liquid assets completely and completely max out their credit cards, 9.6 percent of the US population (about 31 million adults and children) could still not afford the set of goods necessary for the basic needs of life. These estimates are highly uncertain because of weaknesses in the data, and much more research is needed on credit cards as an available resource over a relevant time horizon before any definite conclusion can be reached.

We suggest that our work be considered only as a preliminary, initial investigation of our new conceptual measure. There are many data issues with the CE that make implementation of our measure difficult, and better data are needed to implement what we regard as the best approach to measuring poverty. Further work should result in improved measures of estimated poverty in the United States.

ACKNOWLEDGMENTS The authors would like to thank Silin Huang, Nino Kodua, and Matthew Zahn for research assistance, Thesia Garner for help with constructing in-kind transfers in the CE, and Chris Wimer for help

in constructing historical SIPM poverty rates from the data files provided by the Columbia Center for Poverty and Social Policy. Comments from discussants Kathryn Edin, Diane Schanzenbach, Luke Shaefer, and the editors were also valuable, as well as helpful suggestions from Henry Aaron, Katharine Abraham, Constance Citro, David Johnson, and Jonathan Fisher. We thank Scott Fulford, Kevin Moore, and Joanna Stavins for conversations about credit card data.

References

Attanasio, Orazio P., and Luigi Pistaferri. 2016. "Consumption Inequality." *Journal of Economic Perspectives* 30, no. 2: 3–28.

Bavier, Richard. 2014. "Recent Trends in U.S. Income and Expenditure Poverty." *Journal of Policy Analysis and Management* 33, no. 3: 700–718.

Bee, Adam, Bruce D. Meyer, and James X. Sullivan. 2015. "The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?" In *Improving the Measurement of Consumer Expenditures*, edited by Christopher D. Carroll, Thomas F. Crossley, and John Sabelhaus. Chicago: University of Chicago Press.

Board of Governors of the Federal Reserve System. 2021. *Report on the Economic Well-Being of U.S. Households in 2020*. Washington: Board of Governors of the Federal Reserve System. www.federalreserve.gov/publications/2021-economic-well-being-of-us-households-in-2020-dealing-with-unexpected-expenses.htm.

Citro, Constance F., and Robert T. Michael, eds. 1995. *Measuring Poverty: A New Approach*. Washington: National Academy Press.

Cutler, David M., and Lawrence F. Katz. 1991. "Macroeconomic Performance and the Disadvantaged." *Brookings Papers on Economic Activity,* no. 2: 1–74.

Desilver, Drew. 2021. "As National Eviction Ban Expires, a Look at Who Rents and Who Owns in the U.S." Pew Research Center, August 2. https://www.pewresearch.org/fact-tank/2021/08/02/as-national-eviction-ban-expires-a-look-at-who-rents-and-who-owns-in-the-u-s/.

Dillman, Don A., and Carol C. House, eds. 2013. *Measuring What We Spend: Toward a New Consumption Expenditure Survey*. Washington: National Academies Press.

Fisher, Jonathan, and Bradley L. Hardy. 2022. "Money Matters: Consumption Volatility across the Income Distribution." Washington: Washington Center for Equitable Growth.

Fisher, Jonathan, David S. Johnson, Joseph T. Marchand, Timothy M. Smeeding, and Barbara B. Torrey. 2009. "Identifying the Poorest Older Americans." *Journal of Gerontology* 64B, no. 6: 758–66.

Fisher, Jonathan, David S. Johnson, and Timothy M. Smeeding. 2015. "Inequality of Income *and* Consumption in the U.S.: Measuring the Trends in Inequality from 1984 to 2011 for the Same Individuals." *Review of Income and Wealth* 61, no. 4: 630–50.

Fox, Liana E., and Kalee Burns. 2021. *The Supplemental Poverty Measure: 2020*. Washington: US Census Bureau. https://www.census.gov/library/publications/2021/demo/p60-275.html.

Fox, Liana E., Christopher Wimer, Irwin Garfinkel, Neeraj Kaushal, and Jane Waldfogel. 2015. "Waging War on Poverty: Poverty Trends Using a Historical Supplemental Poverty Measure." *Journal of Policy Analysis and Management* 34, no. 3: 567–92.

Hurst, Erik. 2012. Comment on "Winning the War: Poverty from the Great Society to the Great Recession," by Bruce D. Meyer and James X. Sullivan. *Brookings Papers on Economic Activity*, Fall, 189–93.

Insolera, Nora E., Beth A. Simmert, and David S. Johnson. 2021. "An Overview of Data Comparisons between PSID and Other US Household Surveys." Technical Series Paper 21-02. Ann Arbor, Mich.: Institute for Social Research.

Korenman, Sanders, Dahlia K. Remler, and Rosemary T. Tyson. 2019. "Accounting for the Impact of Medicaid on Child Poverty." Working Paper 25973. Cambridge, Mass.: National Bureau of Economic Research. www.nber.org/papers/w25973.

Meyer, Bruce D., and James X. Sullivan. 2012. "Winning the War: Poverty from the Great Society to the Great Recession." *Brookings Papers on Economic Activity,* Fall, 133–83.

Meyer, Bruce D., and James X. Sullivan. 2019. *Annual Report on US Consumption Poverty: 2018.* Washington: American Enterprise Institute.

Sabelhaus, John, David Johnson, Stephen Ash, David Swanson, Thesia I. Garner, John Greenlees, and Steve Henderson. 2015. "Is the Consumer Expenditure Survey Representative by Income?" In *Improving the Measurement of Consumer Expenditures*, edited by Christopher D. Carroll, Thomas F. Crossley, and John Sabelhaus. Chicago: University of Chicago Press.

Slesnick, Daniel T. 1993. "Gaining Ground: Poverty in the Postwar United States." *Journal of Political Economy* 101, no. 1: 1–38.

Social Metrics Commission. 2020. *Measuring Poverty 2020*. London: Legatum Institute.

US Office of Management and Budget (OMB). 2021. *Final Report of the Interagency Technical Working Group on Evaluating Alternative Measures of Poverty*. Washington.

Wimer, Christopher. 2014. "Consumption and Income Poverty: Conceptual and Methodological Issues in the Measurement of Poverty." Working Paper.

Comments and Discussion

COMMENT BY

KATHRYN J. EDIN and H. LUKE SHAEFER Edin began studying the budgets of low-income single mothers in 1987 because of a single chart from a government publication left open on a table in the University of Chicago library. Figure 1 re-creates that chart—which inspired Edin's dissertation and first book (with Laura Lein), *Making Ends Meet*, published in 1997.

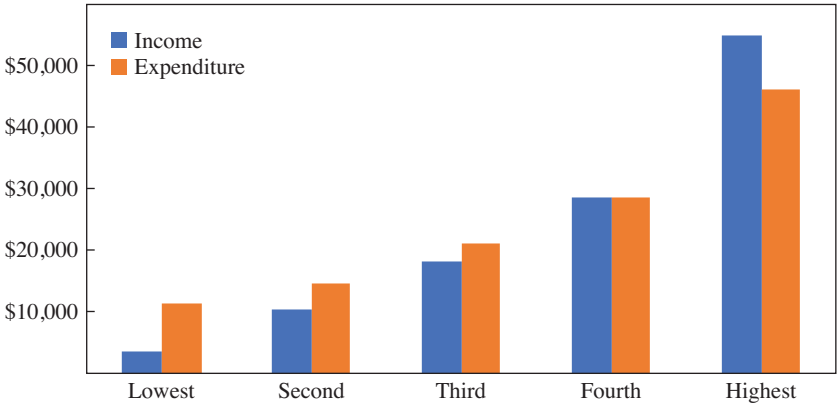
The government publication summarized estimates from the Consumer Expenditure Survey (CE), the same data set Fitzgerald and Moffitt use to construct their supplemental expenditure poverty measure (SEPM). What caught Edin's eye was that in the bottom fifth of the income distribution, households were spending three times as much as they were earning (in fact, expenditures outstripped income for the bottom three quintiles of American households, although to a lesser degree higher up the income ladder).

A footnote to the table attempted to explain the discrepancy, asserting that these households were likely living off past or future income. At the time, Edin was teaching college courses for low-income Chicagoans, many of them receiving welfare, in t he Logan Square and North Lawndale neighborhoods, some of the city's poorest. This explanation did not square with the situations of the students in her class. Most had been struggling economically for many years and thus had no "past income" to draw on, nor, Edin would learn, did they have much access to the credit that would allow them to draw on any "future income."

Figure 1, and the mysteries it implied, sent Edin and anthropologist Laura Lein on a six-year journey across the country to learn more about how low-income, single mothers made ends meet, engaging in multiple in-depth interviews with each respondent while collecting detailed accounts of

Brookings Papers on Economic Activity, Spring 2022: 287–305

Figure 1. Expenditures by Income Quintile, 1986



Source: Consumer Expenditures Survey.
Note: Income is after tax; see US Bureau of Labor Statistics, table 1, https://www.bls.gov/cex/standard/1986/quintile.txt.

income and expenditures. What they found was that even after accounting for all government benefits—welfare plus food stamps, Supplemental Security Income (SSI), and any Earned Income Tax Credit (EITC)—the 214 welfare recipients in their study could only cover three-fifths of their expenses.[1]

This gap between expenditures and income wasn't driven by under-reporting of benefits—the welfare-reliant mothers they interviewed were all collecting benefits from multiple sources and were able to describe what they received from each program in detail. As anyone who simply reviews the benefit levels for most anti-poverty programs will quickly conclude, these programs rarely, if ever, are generous enough to cover a family's core expenses. Instead, Edin and Lein (1997) found that the gap was real. These mothers were scrambling to cover roughly 40 percent of their expenses from other sources. This is because there was a limit to the extent to which low-income families could cut back on their consumption when their formal incomes were insufficient. Those attempting to do so risked losing their children to the state for neglect, as Edin and Lein showed.

How did the 214 welfare-reliant mothers profiled in *Making Ends Meet* bridge the gap between their income and expenditures? Private charities played a role, especially food pantries, which occasionally helped with the bills in addition to groceries. But the most common strategy mothers

1. The study covered some four hundred cases, with the rest relying on low-wage, formal-sector jobs.

engaged in to make ends meet was work. Nearly half of the single mothers interviewed who were welfare-reliant were working at the time of their interviews in order to meet their core expenses, but they hid this work from their caseworkers because their welfare benefits would be cut about a dollar for every dollar earned, leaving them no better off.

Some worked formal jobs under false identities or hopped from job to job to avoid detection. Some were paid under the table. Typical work included hairdressing, childcare, and cleaning homes. A few (9 percent) fenced stolen goods or sold sex to bridge the gap.

Our point here is that when observing discrepancies between income and expenditures there is almost certainly more going on than mere underreporting of benefits. That gap is in part a warning, a red flag that people are likely engaging in survival strategies that may have very real human costs—costs that can compromise the well-being of children and adults alike.

THE HUMAN COSTS OF CONSUMPTION Our more recent research (Edin and Shaefer 2015), conducted two decades later, revealed that little has changed in the years since for poor families scrambling to meet essential expenses. The study that resulted in our book *$2.00 a Day: Living on Almost Nothing in America* took an iterative mixed-methods approach that relied on both large-scale data and qualitative cases.

We followed Paul Heckewelder, from Cleveland, Ohio, for nearly two years. Paul fell into poverty when the family-owned pizza chain failed in the face of the Great Recession. As one location after another closed, nearly all his immediate family lost their jobs. When we first met him at a Cleveland food pantry in 2013, twenty-two people were living in his worn, two-bedroom home. The whole family was relying on Paul's Social Security check. To generate extra cash, Paul and his son scanned the sidewalks of Cleveland's West Side for used air conditioners, aluminum cans, and other metal objects put out for the trash, parts of which could be sold to a recycling center on 65th Street. Additional funds came from sympathetic members of his church: from time to time, a fellow parishioner would put an unmarked envelope full of cash in his hand. These strategies had netted him several thousand dollars in the past year.

Jennifer Hernandez, from Chicago, was living in a homeless shelter with her two children, Kaitlin and Cole, when we first met her in 2012. To generate cash, she collected aluminum cans to sell to the local recycling facility for which she earned about a dollar per hour. She also completed online surveys for modest cash rewards when she could get access to the computers in the shelter's basement.

Jessica Compton, from Johnson City, Tennessee, lived with her husband, Travis, who had had his work hours reduced to zero for nearly two

months when we met them. During that time, the only means of generating cash income for these parents of two girls was for Jessica to sell her blood plasma as often as the law allowed (Travis was barred from donating due to his many tattoos). Just one pound over the weight limit for donating, Jessica took iron supplements in order to pass the tests she was required to take to qualify. Often, she found the experience of donating plasma debilitating. "I get tired. Especially if my iron's down, I get, like, really tired," she told us. In 2019, Americans, most of them low-income, donated more than 50 million units of blood plasma in exchange for cash (Ochoa, Shaefer, and Grogan-Kaylor 2021). At a compensation rate of $30 per unit (our estimate of the going rate), our back-of-the-envelope estimate is that plasma sales generated $1.5 billion in income for low-income Americans that year.[2]

Modonna, who like Jessica was also living in a homeless shelter in Chicago in 2012, could only keep up with her cell phone bill and maintain payments on the storage locker that contained all her worldly possessions due to the contributions of a friend. When she reached the limit for staying at the shelter, she and her teenage daughter moved in with the friend. Just before Christmas, Modonna caught him ogling her daughter. When confronted, he responded by tossing their possessions out the window of the second-story apartment onto the pavement. Then he threw Modonna and her daughter out as well.

For the last few years, we have been studying one county in eastern Kentucky where, in addition to conducting extensive ethnographic observation, we've interviewed more than two dozen community leaders plus about the same number of low-income residents. Of these, three admitted to engaging in under-the-table work like cleaning or babysitting to get by, seven reported gleaning cash contributions from relatives and friends, and seven told us they sold their valuables at pawnshops or on Facebook or sold used clothing and other items on the side of the road. Two admitted to participating in the illegal drug trade, trafficking illicit pain pills and meth, and two said they sometimes sold their prescription drugs—OxyContin and Suboxone—on the black market to get the cash they needed to get by. Four admitted to engaging in illegal cockfighting for money, a popular form of entertainment locally, while several others said they got extra cash from selling their votes at election time, an age-old eastern Kentucky tradition. While each of these survival strategies may have added only modestly to the monthly budget, taken together, they constituted a significant share of

2. We take the roughly 50 million plasma units donated at for-profit plasma centers nationally (Ochoa, Shaefer, and Grogan-Kaylor 2021) and multiply by $30 to reach $1.5 billion.

household income, just as was the case for the 214 welfare-reliant mothers in *Making Ends Meet* more than three decades ago.

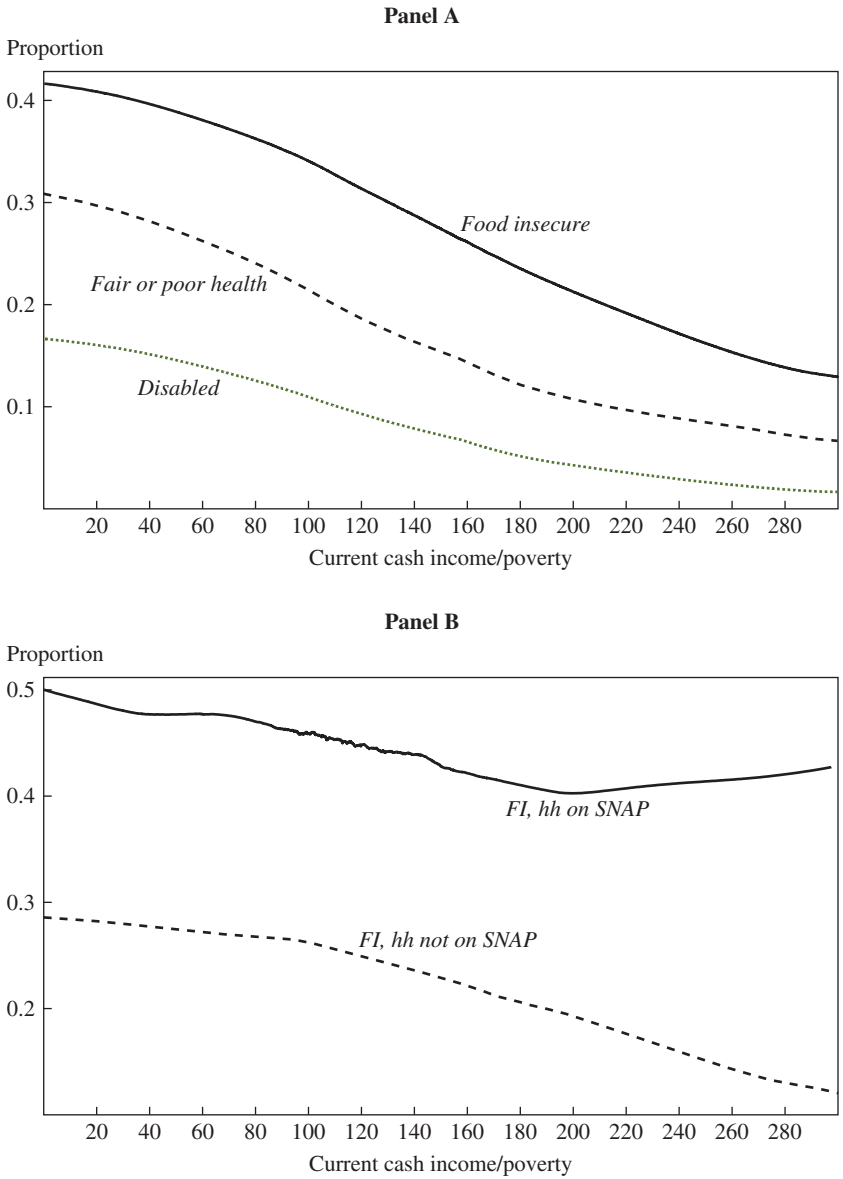
IN POVERTY MEASUREMENT, THINGS ARE NOT ALWAYS AS THEY SEEM As Fitzgerald and Moffitt highlight, the government not only provides cash but also in-kind benefits, such as the Supplemental Nutrition Assistance Program (SNAP) and rental assistance. All else being equal, it is a convention to assume that at any given income level, households receiving in-kind benefits are better off than those who do not receive benefits. Yet through interviews and ethnographic observations (Edin and Shaefer 2015), we found that among families reporting very low cash incomes, receiving SNAP and other in-kind benefits was, in fact, a strong signal of underlying need. Those who could forgo these benefits often had other resources to rely on. Counterintuitive though it seems, we wondered whether families with extremely low cash incomes who were receiving in-kind benefits were actually worse off than their counterparts with similar cash incomes who were not receiving in-kind benefits. This does not mean that SNAP is causing hardship. Rather, SNAP participation is acting as a strong signal of heightened need.

In figure 2 we test this hypothesis using a data set constructed by Anderson, Butcher, and Schanzenbach (2015) linking households across the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC) poverty and food security modules from 2001 to 2011. Anderson, Butcher, and Schanzenbach (2015) engage in one important data cleaning procedure: they drop households with negative income components, typically self-employment or investment losses, who appear to be low-income in the data but who have characteristics far more in line with higher income Americans than with other poor individuals.

Drawing on their data, panel A of figure 2 shows the relationship between annual cash income and three outcomes. As one would expect, households with the lowest cash income are the most likely to be food insecure, report poor or fair health (as opposed to good or excellent), or report a disability. Other analyses reveal they are also the most likely to be Black, less educated, and the least likely to be homeowners (Shaefer and Edin forthcoming).

Panel B is where things get interesting. Among households with children—both households receiving and not receiving SNAP—food insecurity is clearly graded by income. Yet, counterintuitively, at any given income level, those receiving SNAP report far higher rates of food insecurity than those not receiving SNAP. For example, households with annual incomes of 300 percent above the poverty threshold who receive SNAP

Figure 2. The Relationship between Cash Income by Income-to-Poverty Ratio and Well-Being



Source: Shaefer and Edin (forthcoming), in *Social Stratification*, 5th ed., edited by David B. Grusky, Nima Dahir, and Claire Daviss. © Routledge. Reproduced with permission from Taylor & Francis Group.

Note: The underlying data consist of a pooled sample of CPS ASEC modules and food security supplements, 2001–2011, constructed by Anderson, Butcher, and Schanzenbach (2015), who measure income at the household level and engage in one important data cleaning procedure: they drop households with negative income components, typically self-employment or investment losses.

report food insecurity at rates that are demonstrably higher than those not on SNAP who report virtually no cash income.

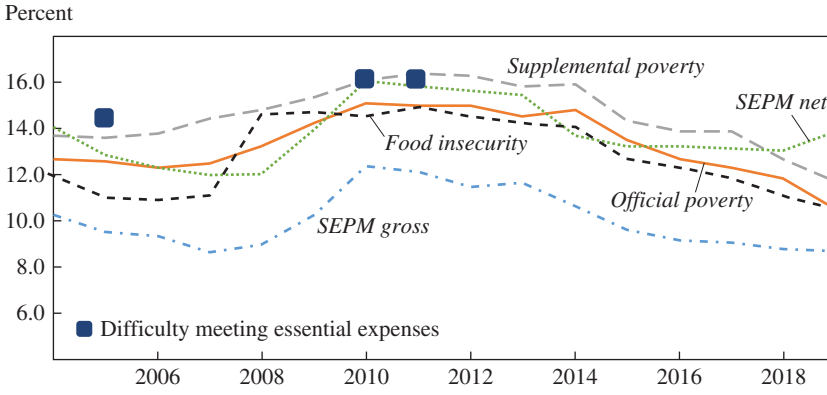
These findings are consistent with an examination by Meyer and others (2021), who find that households reporting extremely low cash incomes but who receive in-kind benefits such as SNAP (much like the families that we profile in our book) experience the very highest rates of material hardship of any group they examined. They conclude that these households "appear to be significantly worse off than the official poor on multiple dimensions of well-being" (8), in line with our findings here and our prior work (Edin and Shaefer 2015). The standard practice of treating all sources of income equally in measuring poverty may inadvertently lead researchers to do a worse job of identifying the neediest households.

A CALL FOR TRIANGULATION Most researchers are keenly aware of the shortcomings of income poverty measures. Here we have illustrated some of the potential shortcomings of expenditure poverty measures as well. In the debate over which measure is the right one—and exactly how an income or expenditure poverty measure should be defined—reaching consensus is made more difficult, if not impossible, by the fact that poverty is often treated as both input and outcome. That is, we very rarely judge measures against anything independent from the internal logic of the measures themselves. Thus, poverty scholars are left to debate, perhaps endlessly, about the right way to account for debt, the proper way to adjust for local cost of living, how to account for in-kind benefits, economies of scale related to family size, and so on. Reasonable people can disagree on almost all these counts, yet many of these decisions have significant implications for our understanding of who is poor and who is not.

Take, for example, the fact that Fitzgerald and Moffitt's SEPM yields more near poor households (households just above the poverty threshold) than their income poverty counterpart, the supplemental income poverty measure (SIPM). They find that if the expenditure poverty line was raised even just slightly, it would lead to 16 million more individuals registering as poor, far more than for comparable income poverty measures. As a result, the decision about where to set the poverty threshold becomes incredibly important in expenditure-based measures. Furthermore, this finding suggests it is reductionist to conclude that expenditure poverty measures yield lower rates of poverty than income-based measures as a general rule.

What is needed is more work to bring the various measures into conversation with one another. Furthermore, researchers should use external outcomes to arbitrate between competing claims by different measures. If what we truly want to do is measure economic well-being, then it makes

Figure 3. Annual Rates of Poverty and Hardship



Sources: Current Population Survey and Survey of Income and Program Participation.

Note: Official poverty measure, supplemental poverty measure, and food insecurity rates are all taken from official government sources. Difficulty meeting essential expenses calculated (at the household level) from the Survey of Income and Program Participation. SEPM rates taken from the paper.

sense to validate and compare these measures based on how well they track with independent, direct measures of well-being, such as material hardship, over time.

Figure 3 offers an external validation of Fitzgerald and Moffitt's measures using plot points provided by the authors. For comparison purposes, we start with two direct measures of well-being: food insecurity, from the CPS food security supplement, and households that report having difficulty meeting essential needs in three comparable years, from the Survey of Income and Program Participation (SIPP). In this figure, we compare these hardship measures with poverty estimates from the official poverty measure (OPM) and Supplemental Poverty Measure (SPM), the SIPM net as estimated by Fitzgerald and Moffitt, and the authors' SEPM gross and net.

Three patterns are evident here. First, though researchers have long grappled with the problems of income-based poverty measures, especially the OPM, the income-based measures track nicely with each other, and with the two hardships in the figure: the line indicating food insecurity and the boxes, which represent difficulty meeting essential expenses. Second, so does the SEPM net. In fact, despite the very different methods used, the correspondence between all these measures in both the level of poverty and trends over time is noteworthy. Could these various measures actually

Table 1. Correlations between Annual Rates of Poverty and Key Indicators of Well-Being, 2004–2019

	Food insecurity	*Unemployment rate*	*Part-time for economic reasons*
OPM	0.9	0.89	0.9
SPM (Census)	0.89	0.85	0.86
SEPM gross	0.75	0.87	0.86
SEPM net	0.61	0.77	0.8

Sources: US Census Bureau, Current Population Survey, and US Bureau of Labor Statistics.

Note: OPM and SPM rates come from official census poverty reports; food insecurity rates from the CPS food insecurity supplement; unemployment rate and part-time for economic reasons from the US Bureau of Labor Statistics.

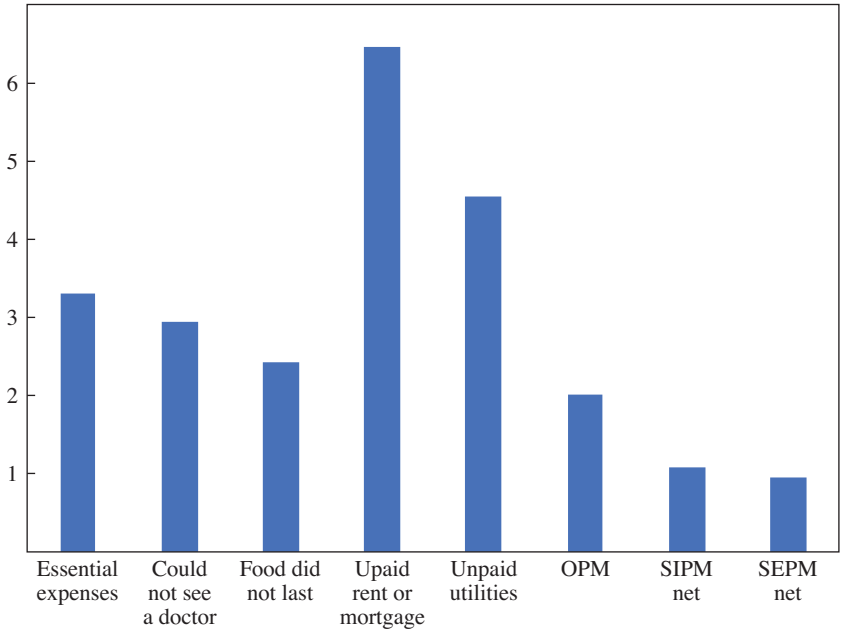
be complementary and give us more confidence that, through triangulation of imperfect measures, we have zeroed in on a basic understanding of how many people are in poverty and how poverty changes over time?

We agree with Fitzgerald and Moffitt that the odd one out here is SEPM gross. It is hard to find face validity in a measure of poverty that is demonstrably below the fraction of US households reporting food insecurity or the fraction reporting difficulty meeting essential expenses, especially when other available measures of income and expenditure poverty suggest such a different story. We can also rule out some potential stories about trends over time. For instance, any measure that would lead to the conclusion that poverty was lower in 2011 or 2012 than in 2004 would have to contend with the fact that multiple poverty measures and direct measures of well-being measures show exactly the opposite.

In table 1, we take this exercise a step further, examining a simple correlation between annual rates of poverty and key indicators of well-being from 2004 to 2019: food security, the unemployment rate, and the percentage of workers who are involuntary employed part-time—imperfect indicators of economic well-being but interesting nonetheless.

We note several patterns here. First, annual rates of all these measures are highly correlated, as the prior figure suggests. Second, the official US Census Bureau income-based measures—especially the OPM—are particularly highly related to the indicators included in table 1. Strikingly, all OPM correlations are at or above 0.89, higher than any other measure. The US Census Bureau's supplemental poverty measure (SPM) also has high correlations with all indicators, reasonably in line with OPM. The associations with the SEPM measures, particularly Fitzgerald and Moffitt's preferred measure of SEPM net, remain strong, but less so than the income-based measures. This presents a conundrum: Why, despite all their flaws, do the income measures—and especially the OPM, which we know to be

Figure 4. Ratio of Hardship and Poverty Rates for Children to Those of the Elderly, 2011



Sources: Survey of Income and Program Participation, US Census Bureau.
Note: Material hardship outcomes taken from wave 9 of the 2008 panel of the SIPP; OPM taken from the census for 2012; net SIPM and SEPM are from the paper.

inadequate—track so well with other measures of well-being? Why doesn't SEPM net do better?

In figure 4, we explore what all these measures tell us about a long-standing question about who is poor, revisiting a familiar comparison between children and the elderly that appears in official census poverty reports. We ask what the various measures suggest about the well-being of each group relative to the other. Bars representing the ratio of child-to-elderly poverty by each of the three poverty measures (leaving out the gross SEPM) are on the right. On the left are several bars that show the ratio of child-to-elderly hardship—a direct measure of well-being—drawn from the SIPP.

All measures included here suggest that hardship among children is much higher than it is among the elderly. These ratios range from 2.4 times higher for children compared to the elderly for "food we bought did not last and we didn't have money to get more" to more than 6.4 higher in the case of unpaid rent or mortgage. The OPM puts that ratio at two-to-one—not too far from "food did not last" but far below the rest. Yet the SIPM net and

the SEPM net show child and elderly poverty at near parity. It is hard for us to reconcile the conclusion that poverty is comparable among children and the elderly when children are in households that report considerably more difficulty paying essential expenses, more trouble affording to see a doctor when they need one, not getting enough food, and more difficulty paying the rent and utilities. Our supposition is that this is driven in part by the challenging question of how to treat out-of-pocket medical expenses, which affect the elderly most. This is a thorny issue. Having high monthly out-of-pocket expenses signals that households have fewer resources for other expenses. Yet a zero may signal two very different situations: no underlying health conditions requiring medical expenses or underlying health conditions without the ability to pay.

In conclusion, while we are encouraged by the SEPM alternative that Fitzgerald and Moffitt offer because we believe it enriches our understanding of poverty in the United States, we encourage poverty researchers to consider two points. First, the gap between expenditures and income likely represents more than mere benefit underreporting. Indeed, it may well be a sign that a household is engaging in survival strategies that might have sharp human costs, with deleterious implications for well-being. Second, rather than simply focusing on the internal validity of a poverty measure, consider the degree to which we might accept that all measures are imperfect, work to bring them into conversation with one another, and validate them with external measures of well-being.

REFERENCES FOR THE EDIN AND SHAEFER COMMENT

Anderson, Patricia M., Kristin F. Butcher, and Diane Whitmore Schanzenbach. 2015. "Changes in Safety Net Use during the Great Recession." *American Economic Review* 105, no. 5: 161–65.

Edin, Kathryn J., and Laura Lein. 1997. *Making Ends Meet: How Single Mothers Survive Welfare and Low-Wage Work*. New York: Russell Sage Foundation.

Edin, Kathryn J., and H. Luke Shaefer. 2015. *$2.00 a Day: Living on Almost Nothing in America*. Boston: Houghton Mifflin Harcourt.

Meyer, Bruce D., Derek Wu, Victoria Mooers, and Carla Medalia. 2021. "The Use and Misuse of Income Data and Extreme Poverty in the United States." *Journal of Labor Economics* 39, no. S1: S5–S58.

Ochoa, Analidis, H. Luke Shaefer, and Andrew Grogan-Kaylor. 2021. "The Interlinkage between Blood Plasma Donation and Poverty." *Journal of Sociology and Social Welfare* 48, no. 2: 56–71.

Shaefer, H. Luke, and Kathryn J. Edin. Forthcoming. "Extreme Poverty among Households with Children since the 1996 Welfare Law." In *Social Stratification*, 5th ed., edited by David B. Grusky, Nima Dahir, and Claire Daviss. New York: Routledge.

COMMENT BY

DIANE SCHANZENBACH It is important to obtain accurate, reliable measures of poverty, to answer questions such as how many people are poor, how poverty rates vary across various population groups, and how poverty rates have changed over time. Measuring poverty is deeply challenging, though, because of a host of thorny conceptual and measurement issues. The picture of poverty can vary dramatically depending on how resources are counted, how the threshold of what it means to be poor is defined, and how these concepts are adjusted over time. Fitzgerald and Moffitt do an admirable job making the case for an expenditure-based poverty measure that builds on the framework of the supplemental poverty measure (SPM), which they distinguish from their approach by referring to it as the supplemental income poverty measure (SIPM).

To understand the context of Fitzgerald and Moffitt's approach, it is necessary to briefly review approaches to income-based poverty measures. The official poverty measure (OPM) was developed in the mid-1960s and compares a household's annual cash resources to a poverty threshold that varies by family size. As students of poverty history may recall, the poverty thresholds were originally developed by Mollie Orshansky from the Social Security Administration, who based her threshold on the price of a minimum-cost food diet, multiplied by three to account for other family expenses (Fisher 1992). That original threshold has been adjusted for inflation since then. There are many well-understood drawbacks to the OPM: it ignores taxes and in-kind benefits such as Supplemental Nutrition Assistance Program (SNAP) payments, it considers all money to be available to count against the poverty threshold, and it sets the same poverty threshold in every state despite widely varying costs of living.

Seeking to improve on these limitations, the SPM, which has been released annually by the US Census Bureau since 2011, takes several different approaches to defining both income and needs. The SPM poverty threshold is anchored to US families' spending on a bundle of food, clothing, shelter, and utilities and adjusted over time using a moving average across years. There is some geographic variation built in as well, with poverty thresholds varying by place of residence and homeowner or rental status. The SPM includes government in-kind benefits like food support programs and housing assistance, accounts for taxes (which can be negative for low-income families), and excludes income that can't be used to purchase the minimum bundle of goods, including work and childcare expenses, as well as out-of-pocket medical expenses. To be sure, while the SPM is an improvement over the OPM, it is nonetheless imperfect. An expert consensus panel

convened by the National Academies of Sciences, Engineering, and Medicine is expected to issue a report in the coming months to recommend improvements to the measure.[1]

Many economists are interested in consumption-based measures of poverty. To the extent that flows of consumption and income diverge, consumption may indeed be preferable to the extent that by tracking poverty we are concerned about the share of families experiencing unmet need for food, housing, medical care, or other necessities. Fitzgerald and Moffitt do a great service to the profession by constructing a consumption measure that is directly comparable to the SPM annually produced by the US Census Bureau. Like the SPM, Fitzgerald and Moffitt's SEPM accounts for taxes and in-kind benefits, excludes certain purchases related to medical care, work expenses, and childcare, and compares income to a poverty threshold anchored to spending on core goods that varies by geography and homeowner or renter status.

The overarching takeaway from their work is that poverty measured by the SEPM is quite similar in both level and trend to income poverty, as shown in figure 5 in the paper. Despite SEPM and SIPM tracking nearly perfectly since 2009, there was some divergence between them in the years leading up to the Great Recession, when consumption poverty was substantially lower than income poverty. I note that there was a similar divergence between poverty and food insecurity during the years prior to the Great Recession. Some of this divergence could be explained by low-income households, especially homeowners, consuming out of assets or on credit in this period, as coauthors and I argue in a recent working paper (Anderson and others 2022).

Importantly, the depth and distribution of poverty look more different across the expenditure and income measures. As shown in figure 4 in the paper, while 4.4 percent of households have income levels less than half of the poverty threshold, the share is substantially smaller—only 1.1 percent—in the spending measure. That is good news and suggests that the share in deep income poverty—whether due to real income fluctuations or measurement error—likely overstates the share of households experiencing deep material deprivation. On the other hand, because there is a substantial mass of households with expenditures just above the poverty threshold, it also changes the share in near poverty (less than 150 percent of poverty) from

1. National Academies of Sciences, Engineering, and Medicine, "Evaluation and Improvements to the Supplemental Poverty Measure," https://www.nationalacademies.org/our-work/evaluation-and-improvements-to-the-supplemental-poverty-measure.

one in four by the income measure to one in three by the spending measure. The sensitivity of the rate of poverty to the threshold matters because there is considerable uncertainty as to where to draw the "correct" poverty line—which, as the authors point out, is ultimately socially determined—and that economic well-being is surely not discontinuous at this threshold.

Poverty rates by subgroup vary somewhat across measures as well. Some of these align with my expectations, such as those with high levels of education have a lower expenditure poverty rate than income poverty rate. But other patterns are more puzzling, such as the much higher rate of expenditure poverty among homeowners without a mortgage.

The authors raise an interesting thought experiment on incorporating potential spending in the poverty measure by adding in liquid potential resources such as savings and unused credit cards. It is useful that they highlight the under-explored questions about the role of precautionary savings and credit in consumption smoothing among low-income families, which if spent could reduce the share of households in poverty but generally does not change the time pattern (as they show in figure 9). At the end of the day, this part of the paper is more speculative than anything. For one, the data are not up to the task of credibly assigning potential spending. More fundamentally, I am skeptical that we would want to define someone as not in poverty if they could have increased their consumption by incurring credit card debt (and the inter-temporal consumption shifting that implies).

There are remaining quibbles about measurement worth noting. A challenge of using spending data is always how to measure the flow value of durable goods. Generally, researchers put substantial thought into imputing flow values of vehicles but throw up their hands when it comes to other large-ticket items such as refrigerators and HVAC that are large enough to shift consumption relative to a poverty threshold. The authors do not solve these long-standing problems, and neither will this discussant. There are challenges well known to the SPM approach as well, including how to accurately define who is in the family unit and how to adjust the poverty threshold for family size, geography, and homeownership status. There are many devils in these details; as the authors show, even small changes to the poverty threshold can substantially shift poverty rates.

Overall, Fitzgerald and Moffitt are to be commended for a serious and thoughtful addition to the measurement of poverty with their SEPM. It gives us another approach to measuring household well-being that is based on expenditures and is comparable to a widely cited income measure. Their work also highlights the sensitivity to the poverty threshold and other details of measurement.

REFERENCES FOR THE SCHANZENBACH COMMENT

Anderson, Patricia M., Kristin F. Butcher, Hilary W. Hoynes, and Diane W. Schanzenbach. 2022. "Food Insecurity and the Business Cycle." Working Paper.

Fisher, Gordon M. 1992. "The Development and History of the Poverty Thresholds." *Social Security Bulletin* 55, no. 4: 3–14.

GENERAL DISCUSSION Bruce Meyer argued that the authors' supplemental expenditure poverty measure (SEPM) has not taken recent research into account, including an interagency report and an American Enterprise Institute report.[1] He also claimed that the authors defined "ability to pay" too arbitrarily. He then wondered about the rationale behind the decision to include unused credit lines but exclude other resources that households potentially have access to—like getting a second mortgage or increasing labor supply—and said that consumers' potential consumption does not measure their revealed preferences.

On one hand, Meyer pointed out, the authors' decision to include the ability to borrow may double- or triple-count consumption. For instance, the metric could track a purchase both when a consumer bought something on credit and again when they pay back the loan. On the other hand, Meyer observed, the SEPM excludes much of the first- and second-largest consumption categories: housing and transportation. Meyer's 2012 work with Jim Sullivan found that three-fourths of those considered in poverty by the supplemental poverty measure (SPM) own a car, and around four in ten own a house.[2] He figured that car and home ownership among the poor is likely higher using the authors' measure since it omits the flow of consumption from owned houses and cars.

Rather than recognizing that poverty thresholds are socially constructed, Meyer suggested that the paper recognize thresholds as arbitrary, as did Mollie Orshansky, the economist who developed the official

1. *Final Report of the Interagency Technical Working Group on Evaluating Alternative Measures of Poverty*, 2021, www.bls.gov/cex/itwg-report.pdf; Richard V. Burkhauser, Kevin C. Corinth, Bruce D. Meyer, Angela Rachidi, Matt Weidinger, and Scott Winship, *Addressing the Shortcomings of the Supplemental Poverty Measure*, American Enterprise Institute, July 2, 2021, www.aei.org/research-products/report/addressing-the-shortcomings-of-the-supplemental-poverty-measure/.

2. Bruce D. Meyer and James X. Sullivan, "Identifying the Disadvantaged: Official Poverty, Consumption Poverty, and the New Supplemental Poverty Measure," *Journal of Economic Perspectives* 26, no. 3 (2012): 111–36.

poverty measure (OPM).[3] When the White House initially decided which poverty threshold to use, Meyer said, they targeted a desired poverty rate rather than a specific basket of goods.[4] Meyer figured that the measurement approach would be more meaningful if it were validated by indicators of well-being. Pointing to the American Enterprise Institute report referred to earlier, Meyer also expressed concern about benchmarking the SEPM to the SPM. For instance, he said that the SPM does not align with hardship as well as the official or consumption metrics do.

Robert Moffitt responded in the virtual conferencing chat and argued that Meyer ignored the two criticisms the authors make of consumption poverty: (1) service flows from housing and vehicles are completely illiquid and can't be used to buy food or pay the rent, and (2) a one-period consumption measure ignores the observation that people allocate resources toward consumption differently across different periods.

Katharine Abraham agreed with much of Meyer's commentary. She went on to comment that although the authors are right that a single-year income or expenditure measurement has drawbacks as an indicator of poverty, their solution is inadequate. She offered, as an example, that if a person borrowed on a credit card to pay for something in period one and then paid it back in period two, the expenditure would be counted twice. Instead of looking at income in a single year or double-counting expenditures, Abraham suggested an alternative would be to consider income averaged over a multiyear period, though that would have its own drawbacks and current data are not well equipped to measure income across several years.

Abraham also questioned the utility of comparing poverty rates across measurements. She drew an analogy to comparisons among different measures of unutilized labor supply. The level of a measure that includes involuntary part-timers and marginally attached workers in addition to the unemployed naturally is higher than the level of a measure that includes only the unemployed. Similarly, level differences are only to be expected in poverty measures that are defined differently. While it is meaningful to compare trends, Abraham suggested that level comparisons are unhelpful.

3. Mollie Orshansky, "Counting the Poor: Another Look at the Poverty Profile," *Social Security Bulletin* 28, no. 1 (1965): 3–29.

4. Gordon M. Fisher, "The Development and History of the Poverty Thresholds," *Social Security Bulletin* 55, no. 4 (1992): 3–14.

Abraham further commented that if forced to choose one poverty measure, it makes sense to consider how it relates to indicators of difficulty or hardship at the household level, which the SEPM may not do well.

John Fitzgerald agreed that relating poverty thresholds to material well-being is important. He also acknowledged that the gap between expenditure-based and income-based poverty may point to important aspects of what people must do to earn money. He held that identifying people who are poor in one measure but not in the other may be a useful strategy to identify people in need.

Addressing Meyer's comment, Fitzgerald reflected that the ideal poverty measure, whether it be consumption- or income-based, depends on its purpose. He defended the idea that it is useful for policymakers to know who is still unable to afford a basic bundle even if they do everything in their power, including maxing out their credit cards and spending their bank account balances. Considering Meyer's critique of the available resources included in the proposed poverty measure, Fitzgerald said that the liquidity of available resources depends on the length of the measurement period. He concluded that the proposed SEPM is a step in the right direction toward identifying whose well-being is compromised in a policy-relevant period.

Moffitt agreed with Abraham's concerns about double-counting expenditures and underscored the difficulty of measuring income in a single year. He also pointed out that many policy analysts think measuring poverty annually is too infrequent, since many people experience short-term crises. He wondered about measuring over different time periods or multiple time periods but concluded that measurement is imperfect regardless.

Diane Schanzenbach clarified that the majority of the analysis conducted does not include potential spending; most of the analysis is closely aligned with the approach taken by Meyer and Sullivan.

Caroline Hoxby questioned household construction. She pointed out that poverty is often measured at the household level, but as Kathryn Edin noted in her discussion, significant household spending happens within a network. Given that these networks are especially complicated for low-income families, Hoxby emphasized the importance of carefully crafting household definitions. A single mom, for example, may spend money from her children's father, who may not live with them.

Edin seconded Hoxby's comment; household compositions are fluid, and people may or may not share resources, regardless of where they physically live. She also wondered about how people are faring and remarked that minimum resource bundles are not generous enough.

Robert Hall mentioned the creation of a new panel administered by the National Academies, which will focus on improving inputs to measurement like those for measuring poverty.[5] He then critiqued the notion that consumption properly measures well-being. Since it is necessary to account for transitory consumption, Hall reflected, consumption cannot be a perfect index.

Henry Aaron remarked that including unused credit card balances in measuring poverty is ineffective because it does not reflect a household's ability to sustain a given level of consumption. He added that poverty measurements are meant to inform population-level trends rather than the status of an individual. To measure population-level poverty, Aaron argued that measuring credit card borrowing is a mistake because it must be repaid. He also noted that this line of reasoning may apply to other measures, like increasing labor supply.

Aaron also commented in the chat that much of the controversy over poverty measures arises because poverty lines are used for eligibility for individual benefits (e.g., Affordable Care Act refundable tax credits) and for the distribution of federal funds among states and other political entities. He argued that it is important to show how alternative poverty measures affect different demographic groups and various geographic entities.

Responding to Aaron, Moffitt affirmed what Hall said; consumption is transitory, which impacts the cross-sectional distribution of poverty. He also recognized that some people borrow and others do not, but drawing the line is difficult. Acknowledging that there are no silver bullet solutions, he concluded that the current treatments are unacceptable. He also responded in the chat to Aaron's comment about differential impacts across demographics, pointing out that more work is needed but that the paper does present one table with such results.

Focusing on linguistics, Justin Wolfers questioned how scholars define and discuss poverty. To Wolfers, the discussion seemed to take a prescriptive approach to defining poverty. He wondered if those participating in the conference are the best people to prescribe what poverty means, since attendees are upper-middle-class, and few have spent much time in poverty. Policymakers or voters holding policymakers accountable, Wolfer continued, may be more relevant consumers of data on poverty than the

5. National Academies of Sciences, Engineering, and Medicine, "An Integrated System of US Household Income, Wealth, and Consumption Data and Statistics for Policy and Research," https://www.nationalacademies.org/our-work/an-integrated-system-of-us-household-income-wealth-and-consumption-statistics-to-inform-policy-and-research.

conference attendees. Addressing the panel, he asked that researchers consider pursuing a more descriptive approach to poverty by considering what people want to know when they ask about poverty.

Luke Shaefer agreed with Wolfers's commentary and reported that his students say they would set the poverty threshold higher than the current thresholds. To Moffitt's point, Shaefer underscored how important shocks are at different positions in the income distribution; a person who is just above the poverty threshold may have a tougher time recovering from a short bout of instability than someone in a different place in the income distribution. He also expressed interest in how the population would set the poverty threshold.

Meyer then said that he doesn't believe a consumption poverty measurement is ideal and that he is in favor of using several measures and cutoffs. He then emphasized an advantage of consumption: it captures revealed preference, which indicates what people feel they can afford.

Fitzgerald argued that some trends are long-term, and some constraints are short-term, so choosing any time frame will be imperfect. Nevertheless, he continued, you must decide how to proceed, and trying different measures is a way of starting that process.

Schanzenbach highlighted that many of the comments were about the challenges to measuring the flow value of durable goods. She proposed being transparent and considering complicated questions, like the household question Hoxby raised. Family size adjustments are also often wrong.

ROBERT J. SHILLER
Yale University and NBER

ANNE K. THOMPSON
Massachusetts Institute of Technology

What Have They Been Thinking? Home Buyer Behavior in Hot and Cold Markets: A Ten-Year Retrospect

ABSTRACT This is an update of a paper that we published with Karl E. Case in *Brookings Papers on Economic Activity* in 2012. The paper analyzes data from our annual questionnaire survey of US home buyers to understand their expectations for future home price changes. We again see a period of rapid price increase as we did in our surveys a decade ago. We find that home buyers were generally well informed, and their short-run expectations were conservative, typically underreacting to the year-to-year changes in actual home prices. Housing bubbles can be seen in their long-term (annualized ten-year) home price expectations. The long boom that preceded the 2007–2009 crisis was associated with changing public understanding of speculative bubbles. During the early years of this decade-long rebound, both short- and long-term expectations were out of line with actual changes in prices. Since 2013, long-term expectations have converged with short-term expectations and actual price changes in most locations, and all three series have moved in synch. With the onset of COVID-19, in 2021 actual and anticipated appreciation diverged once again. This time, however, short-term expectations surged above long-term expectations but remained far below actual appreciation rates. Buyers presumed a coming slowdown in the market that has yet to materialize.

Conflict of Interest Disclosure: Robert J. Shiller works with the Chicago Mercantile Exchange on home price index futures and serves on the S&P CoreLogic Case-Shiller Index Committee of Standard and Poor's S&P Global. He also consults for Barclays Bank. The authors did not receive financial support from any firm or person for this paper or, other than the aforementioned, from any firm or person with a financial or political interest in this paper. Other than the aforementioned, they are currently not an officer, director, or board member of any organization with an interest in this paper.

Brookings Papers on Economic Activity, Spring 2022: 307–344

We report here on a continuation of our questionnaire survey study on the expectations and understandings of real estate markets by the general home buying public in the United States during the early years of the twenty-first century, 2003–2021, encompassing the 2007–2009 financial crisis and the COVID-19 pandemic. We last reported on our study ten years ago in this forum (Case, Shiller, and Thompson 2012). The aim, as before, is to better understand some extreme and surprising movements in home prices.

Our questionnaire survey was first conducted in 1988. It resumed in 2003 and has since been conducted annually at the Yale School of Management. It is the longest-running regular survey of residential real estate expectations. It also differs from all other housing expectations surveys in that it samples home buyers who purchased a house just prior to the survey date rather than public opinion at large. In markets with high costs of trading or other barriers to trading, it may be that market prices reflect the views of those people who are most enthusiastic or active in the market rather than the public at large.

The survey is unusual in that our paper questionnaire invites the participants to offer comments, in their own words, amid our questions. We then count how often they bring up certain concepts or narratives. In this sense, our survey is more like a focus group than most surveys. It allows participants to explain themselves, and for us to quantify their explanations.

We have attempted throughout to keep our questions in everyday language that people commonly use. We do not ask about real prices or real interest rates, even though economists would like to know what people think about such things. The reality apparently is that most people don't think about such things at all. In all of our questionnaires through 2021, only one respondent ever left a comment using the phrase "real price," and not a single respondent left a comment using "real interest rate."

People often sit for years thinking that they should move to a different house or from renting to owning but wait for some stimulus to push them to actually make the change. There is no paradox in seeing maximal price increases over the last year amid only moderate expectations for future long-term price increases. The upswing in home prices is not so improbable if we reflect that these have been disruptive times (with the aftermath of the 2007–2009 financial crisis and then the COVID-19 pandemic). Lives were disrupted by unemployment and by the stress of lockdowns, quarantines, and deaths in the family, which have left people stewing about their living situations. Family stresses have been rife, and these have been matched by rising crime rates and political polarization. These pressures may be

driving some people to find the perfect house with space for meaningful new and different activities, enough to dominate sales of homes. There may soon be currents of change in the housing market caused by rising interest rates. In addition, the impact of the war in Ukraine is beyond the purview of typical econometric forecasting models and thus difficult to predict.

If we really want to understand why home prices have surged, we need to look at the people who paid these prices. We are in just such a home price boom now. Home prices have been soaring lately. The monthly S&P CoreLogic Case-Shiller U.S. National Home Price NSA Index increased by 114 percent in the decade from the February 2012 post–financial crisis bottom to February 2022.[1] In real, CPI inflation–corrected terms, this was a 71.9 percent increase. In one year alone, from August 2020 to August 2021, the national index increased 20 percent. In real terms, this was a 14 percent increase. This was bigger than any year during the housing boom that ended in a crash at the time of the Great Recession or the financial crisis of 2007–2009. According to our data, this was the biggest one-year increase since 1946, when soldiers returning from World War II found an inadequate supply of houses. The beginning of the baby boom was then driving demand for more floor space, but the War Production Board had shut down most housing construction to free up resources for the war. We might call the COVID-19 pandemic a sort of war, but this war is not over and this time there is no baby boom: the birth rate is unusually low. It is also a time when questions are being asked about the suitability of location and layout of existing homes in the post-pandemic economy, questions which some say ought to depress existing home prices.

This housing boom has certainly been widely noted despite the pandemic. Candidate explanations for this boom of course often refer to expansionary monetary policy. The Federal Reserve kept the effective federal funds rate near zero for a record seven years, from 2009 to 2016 and as low as seven basis points at the end of 2011, just before the current housing boom took flight. In March 2020, the Federal Reserve again cut the federal funds rate to near zero, just as the World Health Organization declared COVID-19 a pandemic. Congress then passed the $2.2 trillion Coronavirus Aid, Relief, and Economic Security (CARES) Act. But these factors are not likely thought of as the exclusive causes of the boom. The crisis ultimately began from the people, not stabilization authorities. Their understanding, their motives for action, must also be understood.

1. S&P Dow Jones Indices, "S&P CoreLogic Case-Shiller U.S. National Home Price NSA Index," https://www.spglobal.com/spdji/en/indices/indicators/sp-corelogic-case-shiller-us-national-home-price-nsa-index/#overview.

Turning points in economic series do not correspond closely to major policy announcements. For example, the first-time home buyer tax credit—10 percent of purchase price capped at $7,500—created by the Housing and Economic Recovery Act of 2008 (and expanded to a cap of $8,000 in the American Recovery and Reinvestment Act signed by President Obama in 2009) softened the blow to the housing market in 2010, bringing in new home buyers and boosting prices. However, the expansion was short-lived. It expired in 2010, causing prices to slip again in 2011, before the current housing boom began in 2012. The Federal Reserve's announcement in March 2020 of an extreme program of monetary stimulus and the passage of the CARES Act that same month appear to have arrested the sharp March-April slide in stock prices but do not explain the further rise to new highs by the end of 2021. A number of federal bills, if signed into law, would have directly supported the housing market. A bill entitled the First-Time Homebuyer Act of 2021, which would raise the maximum credit to $15,000, was introduced in the House of Representatives on April 28, 2021. The bill never passed the Senate. The Build Back Better Act of 2021 would have raised the maximum credit to $20,000. The possibility that something like one of these bills, or some state bills, would pass must have encouraged some home buyers. We do see mention of a home buyer tax credit in the comments written by respondents on our questionnaires. However, it was mentioned just seventeen times and limited to the 2010–2012 survey responses.

To help sort through the reasons for the price increase, we turn to our more recent home buyer surveys. The survey asks a random sample of recent home buyers for their thoughts, impressions, and expectations. Our 2012 paper aimed to give better understanding of the path of the housing market before, during, and after the financial crisis of 2007–2009. The objective of this 2022 update is to analyze the perceptions of home buyers over the current decade-long housing recovery.

The natural question then is to explore similarities and differences of the situation in the housing boom leading to the 2007–2009 crisis and now. Fortunately, our surveys carry a lot of information about what people were actually thinking in crisis and noncrisis times. We need to consider their thinking to complement the understanding offered by the literature on the 2007–2009 financial crisis. We noted in our 2012 paper that there have been many theories of the financial crisis and its connection to pricing anomalies in real estate. There are theories that emphasize complacency of lenders about the riskiness of their loans (Mian and Sufi 2009; Demyanyk and Van Hemert 2011) and theories relating to money illusion at a time of changing inflation rates (Brunnermeier and Julliard 2008). Add to these theories about

rating agencies' conflict of interest (Mathis, McAndrews, and Rochet 2009) and theories about failures to regulate the shadow banking system (Gorton 2010). But these theories of complacency or failures to regulate must themselves be understood in terms of changing thinking of the general public.

I. Our Survey of Home Buyers

Our first survey, mailed in the late spring of 1988, consisted of a ten-page questionnaire, which we sent to a random sample of 500 home buyers in each of four locations within metropolitan areas around the country: Alameda County, California (Oakland and much of the East Bay, in the San Francisco-Oakland-Fremont, CA Metropolitan Statistical Area); Middlesex County, Massachusetts (Cambridge and the areas north and west, in the Boston-Cambridge-Quincy, MA-NH Metropolitan Statistical Area); Milwaukee County, Wisconsin (the core of the Milwaukee-Waukesha-West Allis, WI Metropolitan Statistical Area), and Orange County, California (which includes Anaheim and Irvine in the southern part of the Los Angeles-Long Beach-Santa Ana, CA Metropolitan Statistical Area). These four were chosen to represent what were viewed at the time as two "hot" markets (Los Angeles and San Francisco), a "cold" (post-boom) market (Boston), and a relatively stable market (Milwaukee).

Annual surveys, which began in 2003, followed the model of the 1988 survey. The questionnaires were identical except for the names of the local areas across the four survey locations. Participation was limited to those who had actually closed on a home that spring. In a typical year, only about 5 percent of the nationwide housing stock changes hands. Thus, our respondents do not necessarily represent the universe of homeowners, home seekers, or home sellers.

The response rate, shown in table 1, to that 1988 questionnaire survey was strong: of 2,030 surveys mailed, 886, or 43.6 percent, were ultimately completed and tabulated. Case and Shiller (1988) presented the results of that survey and concluded, "While the evidence is circumstantial, and we can only offer conjectures, we see a market largely driven by expectations. People seem to form their expectations from past price movements rather than having any knowledge of fundamentals. This means that housing price booms will persist as home buyers become destabilizing speculators" (45). In addition, we found significant evidence that housing prices were inflexible downward, at least in the absence of severe and prolonged economic decline.

After a gap of fifteen years, we replicated the 1988 survey for the *BPEA* conference in the same four counties. We have repeated the questionnaire

Table 1. Home Buyers Survey Response Rates, 1988–2021

Year	*Surveys returned*	*Response rate (%)*
1988	886	43.6
2003	705	35.3
2004	456	22.8
2005	441	22.1
2006	271	13.6
2007	300	15.0
2008	545	27.3
2009	370	18.5
2010	375	18.8
2011	319	16.0
2012	332	16.6
2013	368	18.4
2014	248	12.4
2015	296	14.8
2016	299	15.0
2017	320	15.9
2018	289	14.5
2019	284	14.2
2020	329	16.5
2021	266	13.3
All years	7,699	19.2

Source: Authors' calculations.

survey in the late spring of each year since then. Except for the addition or deletion of some new questions at the end, the questionnaire has remained almost exactly the same in all surveys. In Case, Shiller, and Thompson (2012), which went to press just as the housing recovery began to take shape, we noted that short-term expectations of home prices had strengthened while long-term expectations had weakened. We concluded "although a recovery may be plausible . . . we do not see any unambiguous indication in our expectations data of a sharp upward turnabout in demand for housing" (293).

We now have completed the process a total of twenty times, and this paper presents a first look at the aggregate results. The response rate has varied over time. It has remained below 20 percent since 2009. In 2014, it reached a low of 12.4 percent. The 2021 response rate was 13.3 percent.

II. Home Buyers Are Knowledgeable about the Latest Year's Actual Price Change

In table 2 we compare the actual behavior of home prices in the four metro areas with what respondents thought was happening in their area at the time. For each metro area across all twenty survey years, we calculated the

Table 2. Correlation between Perceived and Actual Price Trends by Survey Location 2003–2021

	Actual price trends				
Perceived price trend	*Alameda County, CA*	*Middlesex County, MA*	*Milwaukee County, WI*	*Orange County, CA*	*All*
Rising rapidly	0.735	0.783	0.826	0.790	0.749
Falling rapidly	–0.831	–0.647	–0.697	–0.697	–0.729

Source: Authors' calculations.

Note: Results are simple correlations for 2003–2021 between the percentage of respondents in the indicated location who gave the indicated response and the actual percentage change in the S&P CoreLogic Case-Shiller Home Price Index for that metropolitan area (measured from the second quarter of the year before to the second quarter of the survey year). Data for each location and pooled across all twenty survey years.

correlation of the actual year-to-year change in the second-quarter average of the local S&P CoreLogic Case-Shiller home price indexes with the percentage of respondents in the corresponding survey area in that year's survey who said prices were "rising rapidly" and with the percentage who said "falling rapidly." If buyers were well informed, one would expect to see a high positive correlation of the year-over-year price increase with the percentage saying "rising rapidly" and a high but negative correlation with the percentage who said "falling rapidly."

The simple correlation coefficients were high in 2012 when the housing recovery began. The additional ten years of survey data have not altered this relationship. In 2021, the correlation coefficients were close to 2012 levels. These measures are high in all four locations, and all have the correct sign, indicating that respondents' perceptions of actual recent price changes have been largely on target for the past twenty years.

Figure 1 provides more detail. It plots the nominal S&P CoreLogic Case-Shiller home price indexes for all four metro areas since 1987; the tables within each panel report the full breakdown of responses to the question about price trends (question 13 in the questionnaire) and expectations (question 26E) in eight of the annual questionnaire surveys (whose dates are indicated in the figure by vertical bars). In all four locations the responses reflected a reasonable knowledge of what was happening at the time of the survey. There was not always consensus, but there was an extraordinary consistency in the results across time and between metro areas.

In Case, Shiller, and Thompson (2012) we noticed a generally realistic expectation of the next year's price increase in the local county but not always so sober an evaluation of the next ten years. Looking at figure 1, one sees that the home price index is quite smooth over time, in contrast to

Figure 1. S&P CoreLogic Case-Shiller Home Price Indexes, 1987–2021

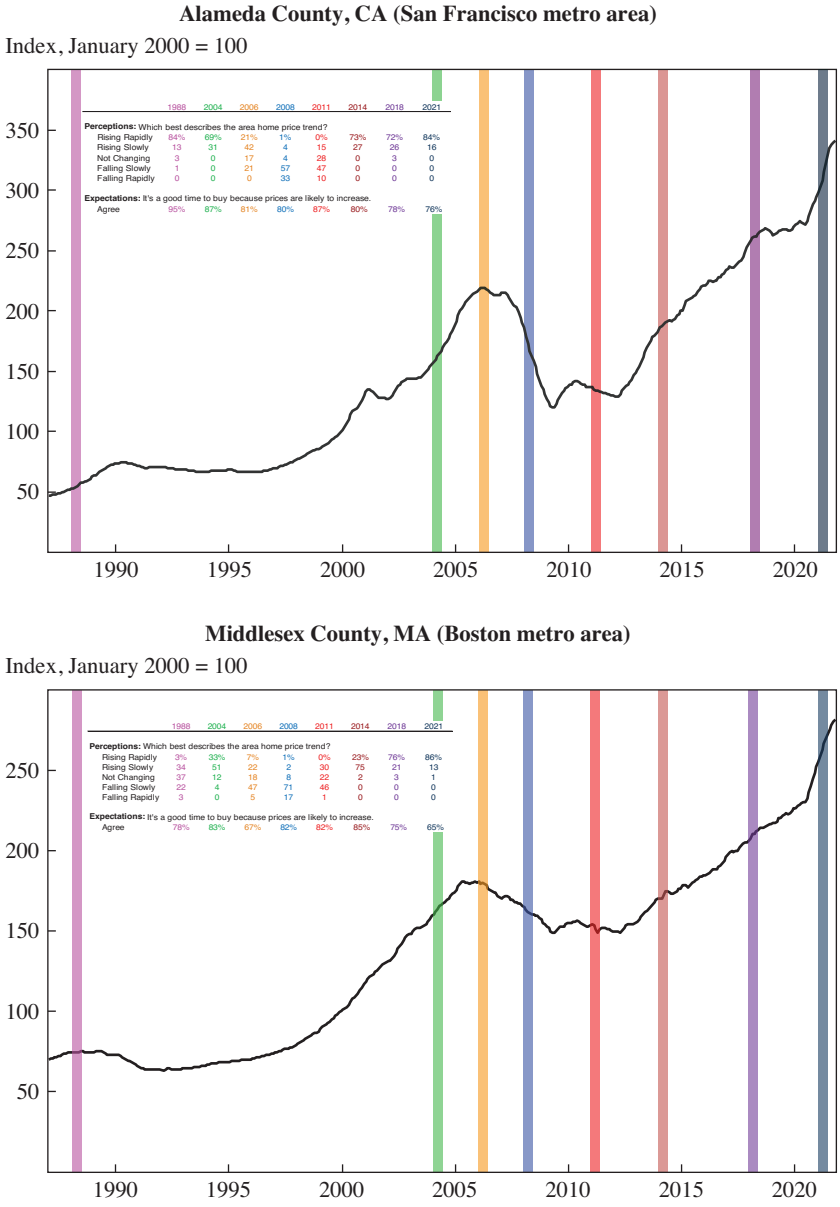
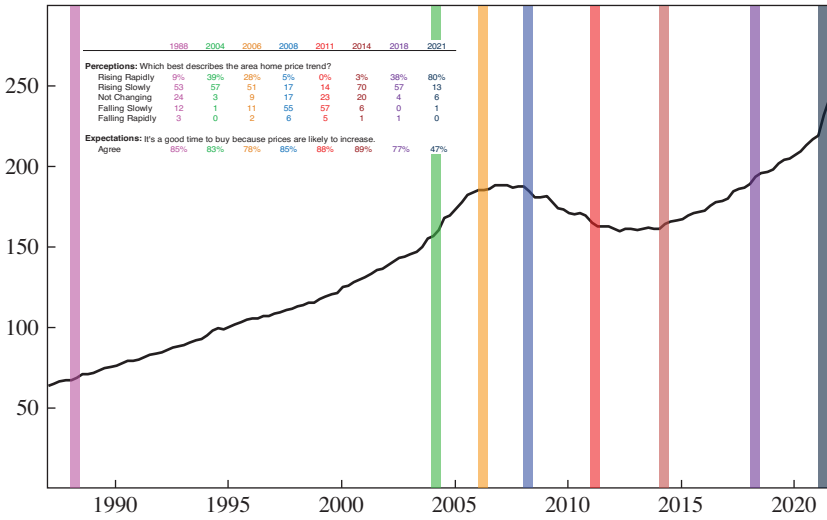


Figure 1. S&P CoreLogic Case-Shiller Home Price Indexes, 1987–2021 (*Continued*)

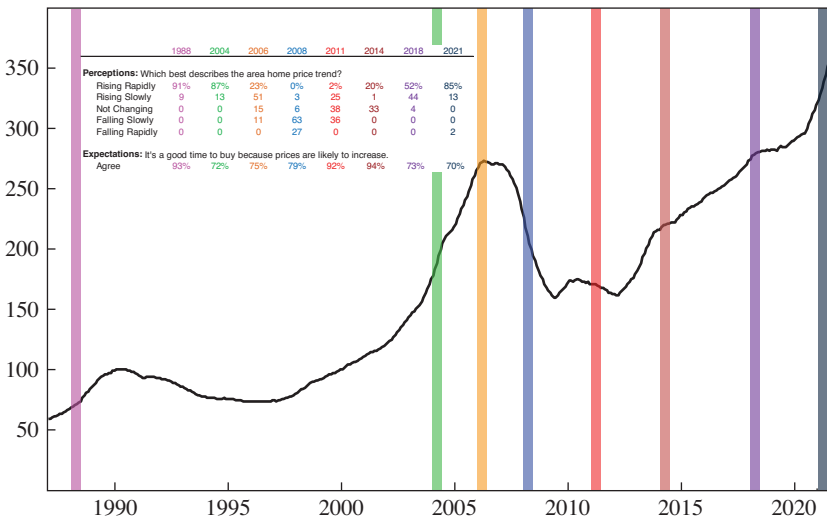
Milwaukee County, WI

Index, January 2000 = 100



Orange County, CA (Los Angeles metro area)

Index, January 2000 = 100



Sources: S&P CoreLogic Case-Shiller and authors' calculations.

Note: Vertical lines indicate quarters in which the home buyer survey perception and expectations are shown. The questions in each table are from survey questions 13 and 26E; the full questionnaire is available at https://www.brookings.edu/bpea-articles/what-have-they-been-thinking-homebuyer-behavior-in-hot-and-cold-markets-a-ten-year-retrospect/.

stock market prices which tend to be very choppy, even from day to day. But longer-term price changes do not look easy to forecast.

From 2015 forward, expected one-year and annualized ten-year home price changes were closely aligned with actual price movement which continued to trend upward. Home buyers were optimistic about price trends. Over 90 percent of respondents in each year from 2015 to 2018 reported prices were trending higher. This was true in Orange County, CA, Alameda County, CA, and Middlesex County, MA. In 2015, 75 percent of Milwaukee County, WI, home buyers viewed prices as moving higher. This share steadily rose over the coming years and by 2018, over 95 percent of buyers in all four counties viewed prices as trending upward.

In 2019, home buyers' optimism began to moderate. Respondents were becoming increasingly apprehensive about another bubble. While 94 percent of Milwaukee home buyers reported prices were trending higher, the shares fell to 88 percent in Middlesex, 74 percent in Alameda, and 68 percent in Orange. These shares remained relatively stable in 2020, the first year of COVID-19. In 2021, however, nearly 100 percent of respondents in each county viewed prices as trending higher.

Figure 1 also shows results from a question asking whether they agreed with the statement, "It's a good time to buy a home because prices are likely to rise in the future"; the vast majority of respondents said yes. On average, through 2020, 85 percent of home buyers agreed with the statement. In every single survey in every county through 2020, the share agreeing with the statement was never less than 67 percent and in most it was over 80 percent. The unanticipated jump in prices in 2021 made buyers a bit more cautious with expectations of future appreciation falling below the previous low in all four counties. Still, the data for Orange (70 percent), Alameda (76 percent), and Middlesex (65 percent) show that buyers remained optimistic. Respondents in Milwaukee were less so, with just 47 percent expecting prices to rise in the future. In all our surveys, the questions about expectations come early in the questionnaire, so that respondents' thinking will not be influenced by narratives explored there.

Question 6 asks respondents how much they think their home is likely to increase or decrease in value over the next twelve months. Question 7 asks on average what they think will happen to the value of their home each year over the next ten years. The wording of these questions has never changed, nor has there been a change in preceding questions, though underlining was added to "on average" and "each year" after 1988, until 2013 when the underlining was omitted. Table 3 tabulates the answers for every year from 2003 through 2021.

Table 3. Short-Term and Long-Term Home Price Expectations, by Survey Location and Year, 2003–2021

	Alameda County, CA	*Middlesex County, MA*	*Milwaukee County, WI*	*Orange County, CA*	*All*
Q6: "How much of a change do you expect there to be in the value of your home over the next 12 months?"					
2003	6.9	4.4	5.5	9.0	6.3
2004	8.4	6.7	5.7	12.4	7.9
2005	9.7	6.4	6.6	8.8	7.7
2006	6.2	1.4	4.8	5.1	3.9
2007	4.7	2.8	6.2	–0.1	3.5
2008	–1.4	–0.6	2.0	–2.3	–0.5
2009	2.2	1.9	1.2	0.6	1.5
2010	3.8	2.2	2.8	3.8	3.0
2011	1.4	1.9	1.2	0.3	1.2
2012	4.4	2.2	2.3	3.6	3.1
2013	8.8	4.4	2.9	7.5	5.7
2014	10.0	4.1	5.3	6.1	5.9
2015	8.0	4.9	3.1	5.3	5.3
2016	5.7	4.2	3.6	6.0	4.7
2017	6.1	5.6	4.5	6.1	5.5
2018	7.0	5.7	5.0	4.6	5.6
2019	5.1	4.6	4.5	2.7	4.3
2020	2.2	3.6	4.5	3.4	3.4
2021	6.1	6.1	7.5	6.4	6.3
*Q7: "*On average *over the next ten years how much do you expect the value of your property to change* each year*?"*					
2003	9.2	6.5	6.1	10.4	7.6
2004	12.7	8.7	8.8	13.3	10.5
2005	10.2	8.3	10.5	10.4	9.6
2006	7.7	7.2	8.7	8.1	7.7
2007	9.1	5.3	7.2	7.6	7.0
2008	7.6	6.4	6.4	9.0	7.3
2009	7.0	5.7	7.7	6.3	6.6
2010	9.8	4.6	6.0	6.0	6.4
2011	6.4	3.8	4.4	7.1	5.2
2012	4.4	3.0	3.2	5.0	3.8
2013	4.9	3.1	3.5	5.4	4.1
2014	7.4	3.8	4.2	7.4	5.3
2015	6.2	3.8	3.0	7.1	4.6
2016	4.7	4.8	3.4	6.0	4.4
2017	5.5	4.3	2.9	6.4	4.5
2018	5.3	3.7	3.4	3.6	3.9
2019	6.2	4.2	3.1	4.8	4.4
2020	3.9	3.1	3.3	4.6	3.7
2021	4.4	4.1	4.0	4.5	4.1

Source: Authors' calculations.

Note: For question 6, means are 10 percent trimmed means, that is, the highest and lowest 5 percent of responses were dropped before calculating the mean. For question 7, values ten times or more from question 6 were set to question 6 values, and 10 percent trimmed means were then calculated. Starting with the 2013 surveys, the words *on average* and *each year* in question 7 on the survey were underlined.

III. Changing Patterns of Short-Term versus Long-Term Price Expectations

The numbers in table 3 are trimmed means, calculated after dropping the top 5 percent and the bottom 5 percent of observations. Prior to trimming, we set any question 7 values that were ten or more times the question 6 response to the question 6 value. This was done to correct for potential misinterpretation of question 7. We then did the trimming because a fair number of responses suggested that the respondent did not understand the question or was simply giving a frivolous answer.[2] To some economists the expectation of price increases in excess of 8 percent per year for ten years, as occurs at least once in each of the four locations, will seem absurd. But when one computes the actual rates of nominal appreciation in the S&P CoreLogic Case-Shiller 10-City Home Price Index (a nationwide measure) from 1996 to 2006, just before the peak, it turns out to be a little above 10 percent per year on average for that ten-year period. Indeed, more than half of our city-specific indexes show ten years of returns averaging in excess of 10 percent per year. This was taking place precisely as the expectations that we are describing in our survey were being formed.

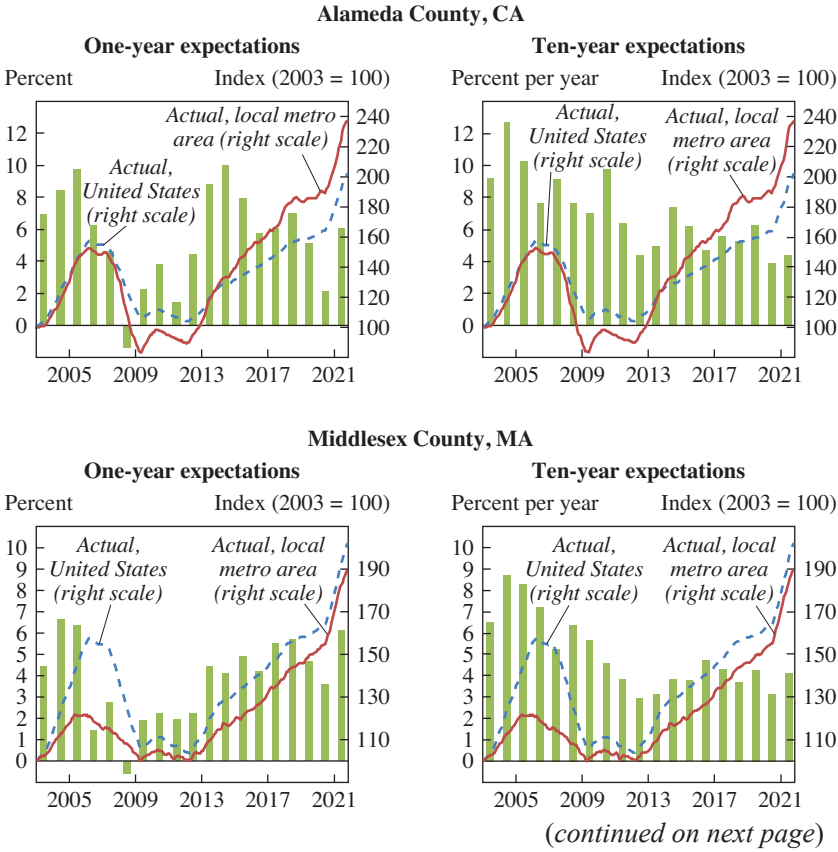
In figure 2 the bars in each of the left-hand panels show, for each year from 2003 to 2021, the trimmed mean of our respondents' one-year expectation for home prices in each of our four survey locations, with the S&P CoreLogic Case-Shiller 10-City Home Price Index shown on the scale to the right. The right-hand panels show the trimmed means of our respondents' annualized ten-year expectations, again by location.

In Case, Shiller, and Thompson (2012), large differences were observed between the one-year and the ten-year expectations. The one-year expectations are much more volatile and at times negative, whereas the ten-year expectations followed a simpler pattern, peaking around 2004 and then only gradually declining.

Both kinds of expectations are important. If one-year expectations are high, home sellers will have an incentive to wait another year to sell while buyers will have an incentive to buy now rather than next year. But when it comes to the decision of whether to buy at all, and comparing the expected rate of return on the investment with the mortgage rate, the longer-term expectations are likely to be more important.

2. For a full discussion, see the appendix to Case, Shiller, and Thompson (2012).

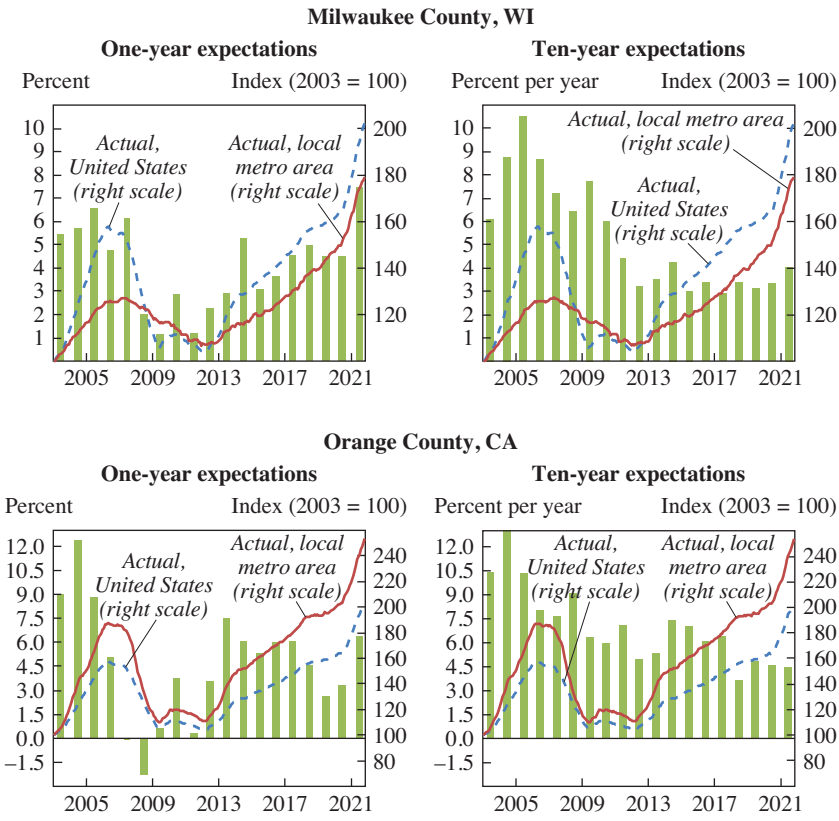
Figure 2. Expected Home Price Growth and Actual Home Prices, 2003–2021



(*continued on next page*)

When we presented our 2012 paper at the *BPEA* conference, discussants questioned whether respondents understood the question about ten-year expectations. They may not be comfortable with defining an average of one-year expectations for ten years. So in our survey the following year, we asked respondents to translate their annualized ten-year expectation to a total ten years' increase. We put an additional question 7B at the very end of the 2013 questionnaire, so as not to change the context of other questions on the questionnaire: "7B. (Clarifying question 7 answer) How much higher do you expect home prices to be, in percentage terms, in 10 years?" In all four counties, 232 respondents answered both question 7 and question 7B. Of these, 22 percent gave the same answer on both, suggesting that they misunderstood the question. The average over the four counties of the question 7B median answer to the question 7 median answer was 6.1. Only

Figure 2. Expected Home Price Growth and Actual Home Prices, 2003–2021 (*Continued*)



Sources: S&P CoreLogic Case-Shiller and authors' calculations.
Note: Bars indicate expected growth (left scale).

one respondent mentioned compounding. This sole respondent said that the ratio should be over 10, because of the power of compounding.

Table 4 presents yet another way of looking at the expectations data. Here we look at expectations since 2003, both short- and long-term, and at actual rates of change in nominal home prices annually from 1996 through 2021 for Orange and Middlesex Counties.

When asked to project how much their home's value would increase or decrease in the next twelve months as well as in each of the following ten years, home buyers in both locations were optimistic. But even these expectations were not unreasonable given the performance of the market before 2006. Price increases in Orange County were actually accelerating

Table 4. Actual versus Expected Short- and Long-Term Home Price Expectations in Orange, CA, and Middlesex, MA, Counties

	Expected annual value increase			
Year	*Next year (%)*	*Annualized next ten years (%)*	*Actual one-year price increase (%)*	*Implied value of a home worth $100,000 in 2000*
Orange County, CA				
2000	n.a.	n.a.	—	100,000
2001	n.a.	n.a.	9.8	109,801
2002	n.a.	n.a.	11.8	122,727
2003	9.0	10.4	18.3	145,130
2004	12.4	13.3	31.2	190,457
2005	8.8	10.4	18.6	225,916
2006	5.1	8.1	15.1	259,942
2007	–0.1	7.6	–3.2	251,605
2008	–2.3	9.0	–24.3	190,505
2009	0.6	6.3	–19.7	153,027
2010	3.8	6.0	8.8	166,465
2011	0.3	7.1	–3.1	161,350
2012	3.6	5.0	–2.2	157,723
2013	7.5	5.4	19.1	187,794
2014	6.1	7.4	12.1	210,556
2015	5.3	7.1	6.0	223,154
2016	6.0	6.0	5.5	235,381
2017	6.1	6.4	5.4	248,123
2018	4.6	3.6	7.7	267,204
2019	2.7	4.8	1.6	271,367
2020	3.4	4.6	3.8	281,685
2021	6.4	4.5	16.8	328,987
Middlesex County, MA				
2000	n.a.	n.a.	—	100,000
2001	n.a.	n.a.	16.4	116,359
2002	n.a.	n.a.	10.7	128,809
2003	4.4	6.5	11.2	143,235
2004	6.7	8.7	9.5	156,846
2005	6.4	8.3	8.4	170,062
2006	1.4	7.2	–1.3	167,824
2007	2.8	5.3	–4.1	160,952
2008	–0.6	6.4	–5.9	151,460
2009	1.9	5.7	–6.9	141,003
2010	2.2	4.6	4.3	147,093
2011	1.9	3.8	–3.3	142,244
2012	2.2	3.0	–0.2	141,985
2013	4.4	3.1	7.3	152,324
2014	4.1	3.8	8.0	164,452
2015	4.9	3.8	2.3	168,186
2016	4.2	4.8	5.2	177,003
2017	5.6	4.3	6.3	188,076
2018	5.7	3.7	6.8	200,798
2019	4.6	4.2	3.7	208,177
2020	3.6	3.1	4.0	216,477
2021	6.1	4.1	17.4	254,061

Sources: S&P CoreLogic Case-Shiller and authors' calculations.

after 2000, and long-term expectations remained solid as long as prices continued to rise. In general, expectations were not as volatile on the upside and less so on the downside. On the upside, they underestimated the magnitude of the increase in 2004 by 19 percent. When prices started falling sharply in 2007 and 2008, buyers continued to expect healthy 56 appreciation each year over the next ten years, and even their one-year expectations resisted the idea that the severe price drops that were already occurring would continue into the next year. They underestimated the scale of the decline by more than 20 percent in 2008. The range of actual price changes from 2003 to 2009 varied from −24.3 percent to 31.2 percent, a 56 percentage point range. The expected one-year (15 percent) and annual ten-year (7 percent) were far lower. Middlesex County observed similar, but less pronounced, differences between actual and expected rates of change.

While actual prices in both counties improved alongside the home buyers' tax credit in 2010, they headed lower once the credit was no longer available. Prices turned the corner again in 2013, rising month-over-month and year-over-year in all four counties. In Orange County, prices rose by double-digit rates in 2013 (19.1 percent) and 2014 (12.1 percent). Home price appreciation was more tempered in Middlesex County, rising 7.3 percent in 2013 and 8.0 percent in 2014. In both counties, short- and long-term expectations fell below actual appreciation in both years and the annualized expected increase in home prices for each of the next ten years fell below the one-year expectations for the first time in 2013.

The rate of growth in home prices in both counties fluctuated between 2015 and 2020. In Orange County, growth varied between 1.6 percent and 7.7 percent. Middlesex County saw appreciation range between 2.3 percent and 6.8 percent during the six-year period. Actual one-year and ten-year and one-year expected home price changes were closely in synch. Prices in all four counties soared in 2021. Orange County prices jumped 16.8 percent from a year earlier and Middlesex County prices surged 17.4 percent. Alameda (18.4 percent) and Milwaukee (14.5 percent) also saw prices jump in 2021. All counties vastly underestimated the change in prices over the coming year. Why such steep price gains occurred in the midst of the COVID-19 pandemic is a question we will try to begin to answer later in this paper.

IV. The Housing Bubble Narrative

Our sample period includes two major turning points in the housing market: the sudden historic end of the housing bubble around 2006 and the ten-year upswing in the market that began in 2012. Understanding these turning

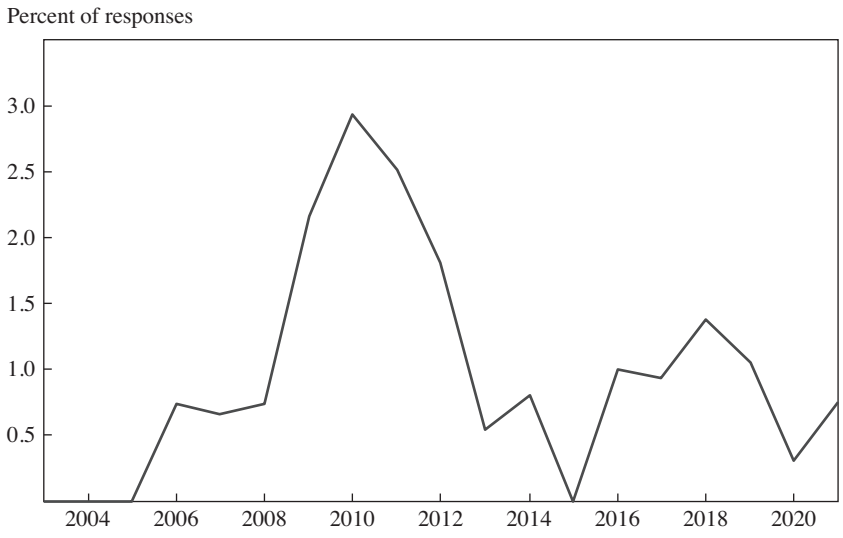
points is central to our objectives. Unfortunately, we observe just two such events in our sample period. But we do have some qualitative information.

Between 2004 and 2007 long-term home price expectations dropped 3.5 percentage points per annum. Ferreira and Gyourko (2011) found that while the beginning of the real estate boom took place at different times in different regions, in all regions the end came at roughly the same time, by 2006. The common themes among respondents in our 2004 survey included a "shortage of houses," a large number of "immigrants," "scarcity of land," "lack of building space," "too many people," and "the desire to have it all." These answers are mostly consistent with perceptions of a shortage of supply. Only occasionally did respondents mention in 2004 that affordability might be an issue. By 2006, the optimistic themes of 2004 were still in evidence but were less prevalent. The most common theme in 2006 was "rising interest rates." Some themes were mentioned repeatedly, in different forms, as suggested by answers such as the following: "high prices," "no equivalent rise in wages," "overvalued homes," "numerous newspapers and media articles speculating on or reporting on slowing sales," and "astronomical price spikes of previous two years simply cannot be sustained."

While the tone of responses went from positive to negative between 2004 and 2007, the opposite occurred between 2009 and 2013. The most common theme in 2009 was the economic downturn and recession, with nearly 25 percent of respondents mentioning this problem. Other common themes in 2009 were "rising unemployment," "foreclosures," "banking crisis," "stock market decline," and "subprime loans." In 2013, the improving economy was mentioned by more than 25 percent of respondents. Other common responses in 2013 included "fewer foreclosures," "rising consumer confidence," "low interest rates," "low inventory," and "investors."

As figure 3 shows, the phrase "housing bubble" did not appear in a single handwritten response in 2004, although one respondent used the term in 2003. By 2006, however, the phrase was being volunteered by a few respondents. As time went on after the crisis, the percentage mentioning "housing bubble" rose until by 2010 over 3 percent of the respondents were volunteering the term. It fell back in 2011 and between 2013 and 2015, "housing bubble" appeared in fewer than 1 percent of responses. The phrase reappeared in 2016 and continued to appear in survey results through 2021. In all of these cases respondents were anticipating a coming bubble, not referring to the financial crisis. The 2010 first-time home buyer tax credit was frequently mentioned that year. Alongside this credit there was a sharp increase

Figure 3. Appearance of the Housing Bubble in Home Buyers Survey Responses, 2003–2021



Source: Authors' calculations.
Note: Share of respondents who used the phrase "housing bubble" anywhere in their answers to the home buyer survey.

in the percentage of respondents that mentioned they preferred not to rent. This share has remained above 15 percent since that point.

V. Short-Term Rationality

The rationality of expectations and the relationship between expected home price changes and realized changes is one of the core applications of our survey. We have compiled an additional decade of data since Case, Shiller, and Thompson (2012) and find that this longer sample reinforces our findings that there is a tendency to underreact rather than overreact. This is significant given the starkly different economic conditions during the first decade of the survey relative to the past decade. We describe the nature of the tests and the impact of the additional data in this section.

To test the rationality of expectations, we need only regress actual home price changes on the expected changes. If expectations are rational, the coefficient of expected changes should be close to one and the constant term zero. With our present data set we can do this only for the one-year expectations, since we have limited ten years of subsequent price data. The

Table 5. Regressions Testing for Rational Expectations of One-Year Change in Home Prices

	Survey location				
	Alameda County, CA	*Middlesex County, MA*	*Milwaukee County, WI*	*Orange County, CA*	*All*
Using S&P/CoreLogic/Case-Shiller Home Price indexes					
Constant	–4.64	–2.39	–2.24	–6.68	–4.87
	(6.41)	(2.91)	(2.88)	(3.73)	(1.94)
Trimmed-mean own-city Expected twelve-month change (Q6)	1.84	1.57	1.32	2.64	2.07
	(1.02)	(0.70)	(0.67)	(0.64)	(0.37)
N	18	18	18	18	72
R^2	0.17	0.24	0.19	0.51	0.31
Using FHFA home price data					
Constant	–2.07	–1.87	–2.09	–5.88	–3.91
	(4.18)	(2.87)	(3.40)	(3.45)	(1.62)
Trimmed-mean own-city Expected twelve-month change (Q6)	1.38	1.41	1.31	2.56	1.89
	(0.67)	(0.70)	(0.79)	(0.59)	(0.31)
N	18	18	18	18	72
R^2	0.21	0.20	0.15	0.54	0.34

Sources: S&P CoreLogic Case-Shiller, FHFA Home Price Indices, and authors' calculations.

Note: Each column in each panel reports results of a single regression for a location. The dependent variable is the actual percentage home price change in the city from the second quarter of the year to the second quarter of the following (future) year. The independent variable is the expected future twelve-month price change (10 percent trimmed mean) from our surveys in the current year. Standard errors are shown in parentheses. All FHFA transactions were used in the 2012 version of this paper. The FHFA index here is limited to purchases-only transactions.

majority of the surveys in each year were returned in the second quarter, so we calculated for the dependent variable the actual price change in each metro area as the percentage change in the S&P CoreLogic Case-Shiller home price index for that area from one second quarter to the next.

Table 5 reports the results. In all four survey locations the slope coefficients are statistically significant and have the right sign, but they are always greater than one. This may be interpreted as implying that homeowners had information that was relevant to the forecast but were not aggressive enough in their forecasts. While the significance and sign of our results here concur with the results in Case, Shiller, and Thompson (2012), there are differences in magnitude. The previous R^2 values were much higher across all locations and the constants and slopes were lower. This can be explained in part by the reduction in outliers over the past decade.

Table 5 shows that home buyers were not overreacting to information with their one-year expectations, but rather underreacting to it. However, this is not necessarily inconsistent with the presence of a bubble. Certainly, the longer-term expectations, whose rationality is harder to judge, seem likely to have been more in line with information in the early years of our sample when they were predicting appreciation of over 8 percent a year for the next ten years.

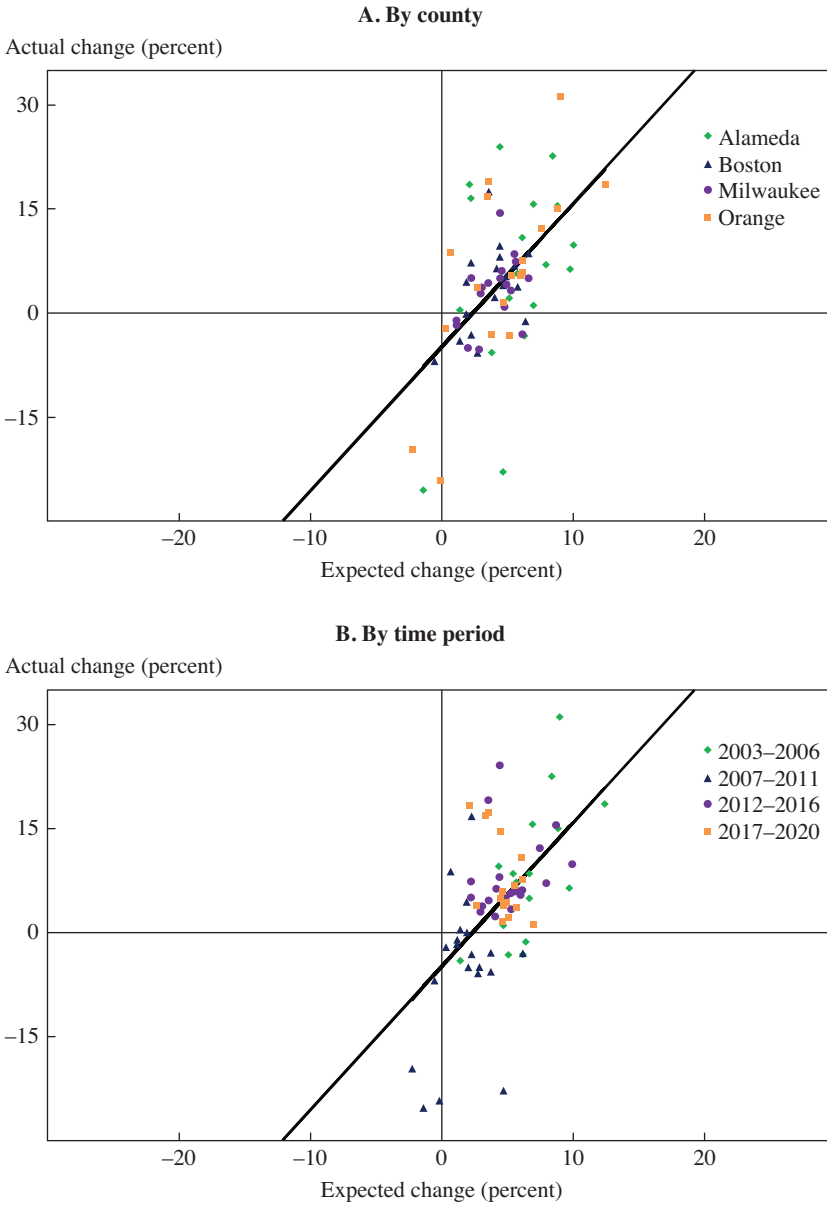
These results do not depend on using the S&P CoreLogic Case-Shiller home price indexes to measure actual price changes. Substituting the home price indexes of the Federal Housing Finance Agency (FHFA, formerly the Office of Federal Housing Enterprise Oversight) purchase-only index yields rather similar results, shown in table 5.

Scatter diagrams of actual against expected one-year price changes for the four counties (figure 4, panel A) and four different time periods (figure 4, panel B) convey how far individuals misjudged the full magnitude of home price movements. This is true both on the downside and the upside. While individuals in all areas underestimated price movements, Alameda and Orange Counties stand out in terms of the scale of misestimation.

Examining these relationships over time reveals that the degree of miscalculation on the upside was largest from 2003 to 2006, when actual price changes greatly exceeded expectations. Not surprisingly, on the downside this occurred between 2007 and 2011. Although prices were plummeting, few expected them to decline at all, and certainly not to the degree that they did. The outliers from 2012 to 2016 occurred primarily in 2012 and 2013. Respondents were cautious coming out of the market collapse but prices rebounded, especially in Alameda and Orange Counties. The 2017 to 2021 extremes all occurred over the last year. The double-digit 2020–2021 jump in home prices that occurred during the pandemic across all four metro areas and nationwide came as a surprise to almost all of our respondents.

Further testing of the rational expectations hypothesis is possible by adding to the regression other information variables available to home buyers when their expectations were recorded. These other variables should have a coefficient of zero if their expectations were rational. We tried two such variables: the actual lagged twelve-month price change in the same metro area and the actual lagged twelve-month price change for the United States as a whole, as measured by the S&P CoreLogic Case-Shiller 10-City Home Price Index. Rational expectations would imply that the coefficient of the one-year expectation should remain at one and the other variables and constant term should be zero. As table 6 reports, both of these variables are insignificant. This is consistent with the rational expectations hypothesis for

Figure 4. Expected versus Actual One-Year Change in Home Prices, 2003–2020



Sources: S&P CoreLogic Case-Shiller, and authors' calculations.

Note: Each observation represents one of the four survey locations in a single year. Actual change shows actual change in metro area home prices from the second quarter of the survey year to the second quarter of the next year. Expected change shows trimmed mean of respondents' expected change in home prices for the next year.

Table 6. Regressions Testing for Rational Expectations of the One-Year Change in Home Prices with Additional Information Variables

Independent variable	*All cities*
Constant	–5.15
	(2.78)
Own-metropolitan area twelve-month price change (%)	2.17
	(0.73)
Lagged own-metro twelve-month price change (%)	–0.03
	(0.23)
Lagged national (ten-city) actual twelve-month price change (%)	–0.01
	(0.21)
N	72
R^2	0.30

Sources: S&P CoreLogic Case-Shiller 10-City Home Price Index and authors' calculations.

Note: The dependent variable is the percentage home price change in the city from the second quarter of the year to the second quarter of the following (future) year. The first independent variable is the expected future twelve-month price change from our surveys, the second is past actual annual price change from the same city, and the third is the past annual US national home price change. Standard errors are in parentheses. The own-metropolitan area twelve-month price change is from trimmed mean of responses to question 6 of the home buyers survey.

Table 7. Regression of Expected One-Year Change in Home Prices on Lagged Actual Price Changes, 2003–2021

	Survey location				
Independent variable	*Alameda County, CA*	*Middlesex County, MA*	*Milwaukee County, WI*	*Orange County, CA*	*All*
Constant	4.53	2.85	3.25	3.20	3.49
	(0.44)	(0.31)	(0.27)	(0.37)	(0.18)
Lagged own-city actual twelve-month Home price change (%)	0.18	0.25	0.27	0.24	0.22
	(0.03)	(0.04)	(0.04)	(0.03)	(0.02)
N	19	19	19	19	76
R^2	0.67	0.67	0.71	0.84	0.72

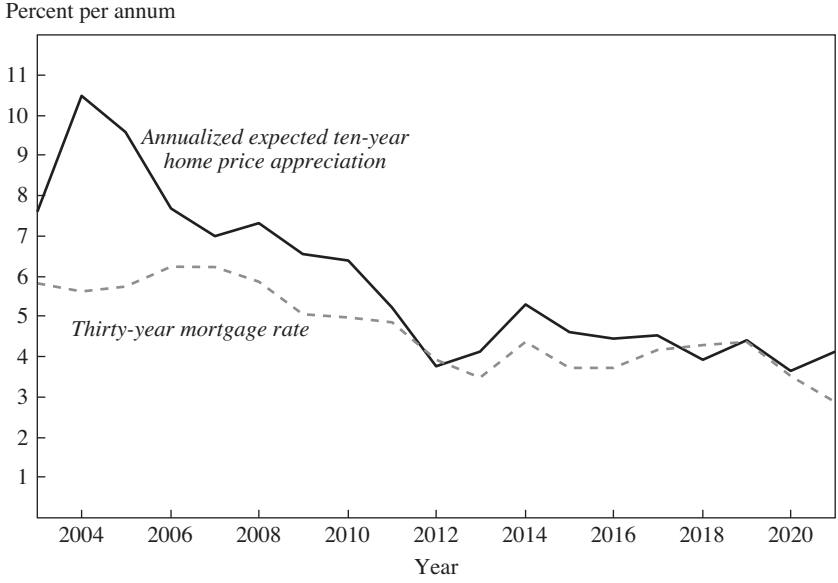
Sources: S&P CoreLogic Case-Shiller Home Price Index and authors' calculations.

Note: Each column reports results of a single regression. The dependent variable is the trimmed mean of the expected one-year change in home values in the indicated location. Standard errors are in parentheses.

the one-year forecasts: respondents are not missing this other information in making their forecasts.

Table 7 reports results of regressions in which the actual and expected price changes switch sides in the equation and the time lag is reversed: we regress the one-year expectation on the lagged actual one-year price change. This allows us to see whether there is a simple structure to expectations. In these regressions R^2 is substantial and in line with our previous

Figure 5. Ten-Year Annualized Home Price Expectations and Thirty-Year Mortgage Rate, 2003–2021



Sources: Authors' calculations and Freddie Mac Primary Mortgage Market Survey.
Note: The annualized ten-year expectation is trimmed mean of responses to the survey question. Average of trimmed means for all survey respondents.

work, ranging between 0.67 and 0.84. Of course, the slope coefficient is less than one because, as we have noted, expectations are less volatile than actual price changes.

Thus, the one-year expectations are fairly well described as attenuated versions of lagged actual one-year price changes, and yet we know from table 6 that they also contain significant information about future price changes beyond what is contained in the lagged actual price change. This conclusion does not mean, however, that any story of feedback in determining price should be modeled in rational terms. Long-term expectations also matter importantly for demand for housing because, as previously noted, they are important to people's decision about whether to buy a home at all.

In figure 5 we see annualized ten-year expectations of home price appreciation from our survey, averaged across our four locations, along with the national average thirty-year mortgage rate, from 2003 to 2021. These expectations, if they could have been trusted, implied enormous profit opportunities in buying a home around 2004: the spread between the two series

was roughly 5 percentage points. Given these expectations, it is no wonder why there was irrational exuberance in the housing market then.

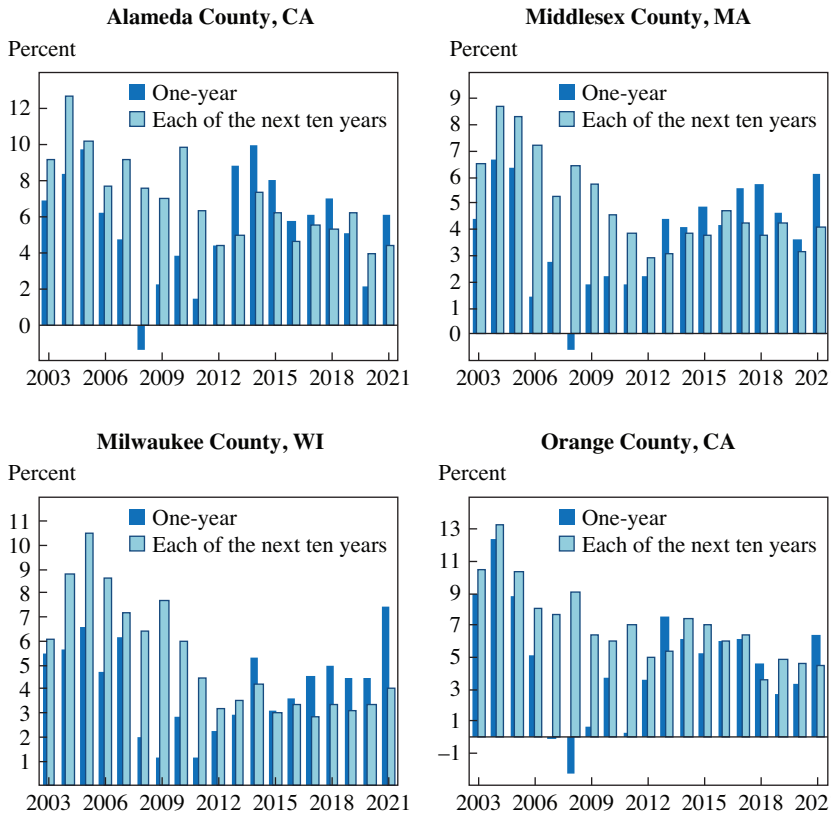
After 2004, however, long-term expectations fell faster than mortgage rates. The apparent profit opportunity narrowed, sharply at first and then more gradually. Neither monetary stimulus nor the other policy measures applied in the wake of the financial crisis succeeded in lowering mortgage interest rates by anything like the decline in expectations.

By 2012 long-term expectations had fallen to a level practically equal to the mortgage rate. Home buyers no longer perceived a long-term profit opportunity by borrowing and investing in a home. This has not changed much over the past decade. Both long-term expectations for home price increases and mortgage rates have been relatively stable. A survey of professional forecasters conducted by Pulsenomics LLC suggests that these professionals became less optimistic than our respondents. Their average expectation for annual home price appreciation for 2012–2016, reported in the June 2012 Pulsenomics survey, was 1.94 percent, about half the ten-year expectation of the home buyers in our 2012 survey. Their average expectation for annual home price appreciation for 2014–2018 in their fourth-quarter 2014 survey was 3.64 percent, closer to, but still below, the ten-year expectation of home buyers in our 2014 survey. The fourth-quarter 2018 Pulsenomics survey included average annual expected price change through 2023. In 2020 and 2021 the average was 2.48 percent, below the annualized 3.9 percent rate predicted by home buyers and far lower than the actual change in the S&P CoreLogic Case-Shiller 10-City Home Price Index, where annual growth averaged 10.5 percent in 2020 and 2021.

Figure 6 clearly shows changes in short-term and annual long-term expectations over the past twenty years. During the housing boom between 2003 and 2006, home buyers' long-term expectations were unrealistically optimistic across all markets. Home prices were soaring, and buyers projected the rate of growth to accelerate in the coming years. When the market took a turn for the worse in 2007, expectations of short-term price growth moderated and turned negative in some metro areas in both 2007 and 2008. Long-term expectations became more unrealistic, with the gap between short- and long-term projections widening significantly. This made sense to some degree, as few would have purchased a home if they expected depreciation. However, the view that the slowdown would be short-lived and prices would rebound significantly and continue to surge was irrational.

The steep drop in prices that occurred between 2007 and 2012 appeared to tame home buyers' expectations. When prices began to rebound in 2012, home buyers were more cautious about both short- and long-term rates

Figure 6. Expected One-Year and Annualized Ten-Year Value Change by County



Source: Authors' calculations.

of appreciation. From 2012–2020, short-term price expectations remained largely below the levels leading up to the crisis across all four metro areas. Even more striking is the steep drop in long-term expectations. Not only have long-term expectations largely remained below 2003–2011 levels over the past decade, but they have fallen below the one-year expectations in 2021.

VI. COVID-19 Takes Center Stage: 2020

The first confirmed case of the novel coronavirus in the United States was reported on January 20, 2020. The virus spread in the United States and abroad, and on March 13, 2020, the US government declared COVID-19 a national emergency. By summer the epidemic looked much worse.

Our 2020 home buyer survey was sent out in early July to those who closed on a home in the first quarter of the year. Thus, most of those surveyed had purchased their home before the national emergency was declared, but after a huge amount of attention was paid to the coronavirus in the media.

A Google Trends search for "coronavirus" shows a sharp peak in March 2020. This was a panic time, with a 33.9 percent drop in the S&P 500 in just over a month from February 19 to March 23, 2020, and an increase in the unemployment rate to 14.7 percent in April 2020, the highest since the Great Depression. A search of ProQuest News and Newspapers shows that the phrase "since the Great Depression" was used 3,368 times in April and May 2020 alone, invariably making a comparison between then and now of unemployment or other indicators of the economy. The newspapers' reporting of these comparisons appears to be motivated by a desire to give a dramatic interpretation to current events, when in fact the Great Depression and the pandemic were really two very different things.

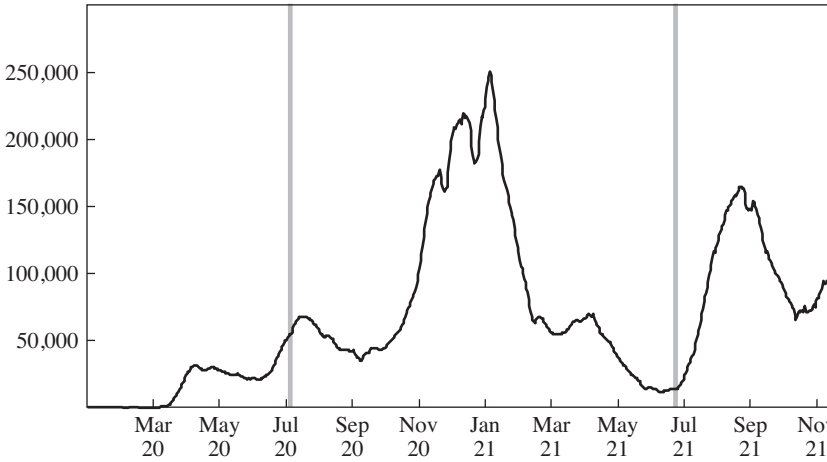
On the very day of the end of the stock market debacle—March 23, 2020—the Federal Open Market Committee of the Federal Reserve came to the rescue with an announcement of aggressive steps that would be taken to stimulate the economy. The announcement, which even said measures would be taken "in the amounts needed to support smooth market functioning," called to mind some famous words of Mario Draghi, of the European Central Bank. During the European Union's euro crisis on July 26, 2012, Draghi said that the bank would do "whatever it takes" to save the euro (Draghi 2012). Newspaper reports on this statement were numerous.

Figure 7 shows the count of COVID-19 cases and the weeks the surveys were mailed. In 2020, COVID-19 cases were rising sharply prior to the mailing. By summer, just as our respondents were completing the survey, fear and uncertainty about the pandemic's full impact was intense. Due to a new wave of virus cases across the country, we chose to add a few questions at the end of our survey in 2020 to gauge whether the spreading coronavirus had altered respondents' perspectives.

Public attention, figure 8, to the coronavirus does not correspond closely to the actual path of the epidemic. A Google Trends search of "coronavirus" or "COVID" shows continued strength of the narrative not closely related to waves in the counts of new cases. There were separate waves of public attention to the coronavirus, following their own epidemic curves, contagion of the narrative only intermittently supported by waves of actual COVID-19 cases.

A similar ProQuest search of US newspapers, blogs, podcasts, and websites showed a similar pattern. In April 2020, nearly 50 percent of these

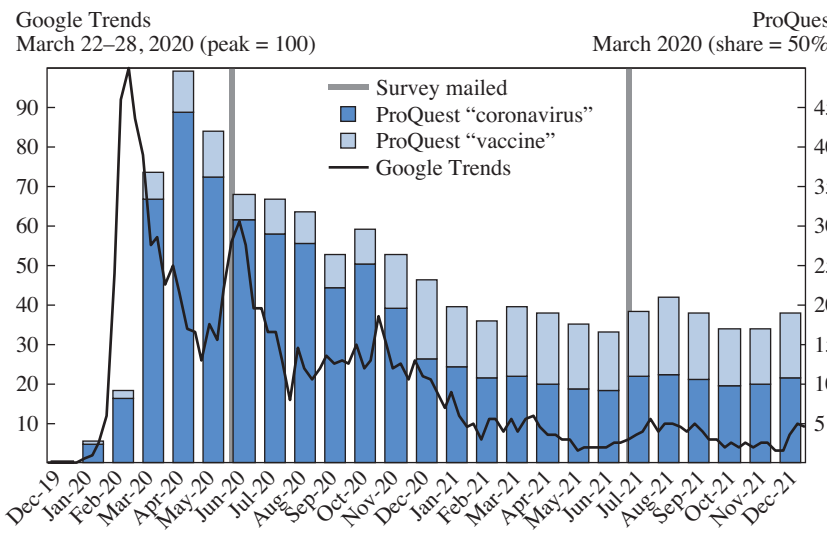
Figure 7. Number of Covid-19 Cases and Survey Mailing Dates



Source: US Centers for Disease Control and Prevention, "Daily Trends in Number of COVID-19 Cases in the United States Reported to CDC," COVID Data Tracker, https://covid.cdc.gov/covid-data-tracker/#trends_dailycases.

Note: Seven-day moving average.

Figure 8. Google Trends Index of Searches and ProQuest Share of Searches for "Coronavirus" or "COVID"



Sources: Google Trends and ProQuest.

Note: Google Trends, "Explore," https://trends.google.com/trends/explore?date=all&geo=US&q=coronavirus%20or%20covid-19. ProQuest (www.proquest.com), using search terms coronavirus* or covid*, and vaccine*.

media articles mentioned COVID-19 or coronavirus. A separate search with the addition of "vaccine" showed that there were relatively few early mentions of vaccines, but by the end of 2020 vaccines were discussed in over 40 percent of these articles.

Responses to the 2020 survey indicate that COVID-19 had certainly affected buyers' outlook, and the major theme was trepidation about the impact the pandemic would have on the housing market and the economy. This was evident not only in the questions added in 2020, but in significant changes in responses to previously asked questions. When asked if any event had changed the trend in home prices over the past two years, 34 percent included the words "covid," "coronavirus," or "pandemic." However, while COVID-19 was frequently mentioned, there wasn't a consensus on the type of impact it would have. Some expected it to drive prices up, but the majority anticipated it would lead prices to fall. A separate question asked respondents what was behind what was going on in terms of recent changes in home prices. COVID-19 was again a common response with 15 percent mentioning it. Other common themes in both questions were low mortgage or low interest rates, shortage of supply, and high demand. Respondents were also asked what they thought would cause current trends to stop. While some were skeptical that the pandemic would continue to hurt the economy, many were optimistic that the development of a vaccine would help the economy to recover.

The 3.4 percent expected one-year change in home values reported in our 2020 survey was the lowest since 2012. These expectations were significantly below the 19.8 percent increase in the S&P CoreLogic Case-Shiller US National Home Price NSA Index over the year ending in July 2021. Between the second quarter of 2020 and the second quarter of 2021, prices shot up 18.4 percent in Alameda County. Survey respondents in Alameda were also the least optimistic, anticipating a 2.2 percent gain in prices over the year, merely 11 percent of the actual price change. In Middlesex (17.4 percent versus 3.6 percent) and Orange (16.8 percent versus 3.4 percent), price gains were five times expectations. Milwaukee County home buyers projected prices would increase 4.5 percent over the year, compared to the 14.5 percent increase that occurred. While the 3.6 percent ten-year annualized expected appreciation for all counties was not as low as the one-year expectation, it was the lowest reported since the start of the survey.

When asked in 2020 whether their "outlook on the economy has worsened since I/we purchased this home," 55 percent of home buyers answered yes. Their view of COVID-19's impact on the housing market was less severe, with 15 percent responding yes to their "expectations for the housing market

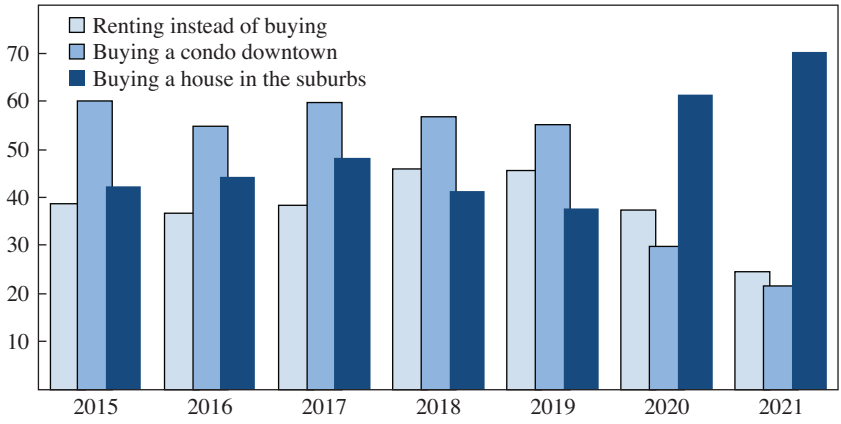
have worsened since I/we purchased this home." Answers to many other questions reveal the fear that home buyers were experiencing in the early days of COVID-19. While prices had tracked steadily higher since 2012, an eight-year low of just 36 percent agreed with the statement "Housing prices are booming; unless I buy now, I won't be able to afford a home later."

Another eight-year low of just 26 percent of respondents perceived home prices as rising rapidly recently. They also kept a close eye on information sources to help determine the price they were willing to pay. A record 92 percent of buyers relied on the internet, Multiple Listing Services (MLS), and newspapers to decide on their offer price.

These home buyers also expected the impact of the coronavirus on the economy to be long-lasting. Nearly 69 percent expected it to continue beyond fall 2020, and 89 percent believed the impact would persist for two or more years. They had purchased a home before COVID-19 had taken hold in the United States, and many had closed on their homes before the first case of the virus was reported here. They received the home buyer survey just as the economy was tumbling into recession. Over a third expected the recession would wreak havoc on the economy. More than 3 million COVID-19 cases in the United States had been reported, and fatalities were increasing. Businesses were closing and mass layoffs were taking place. The unemployment rate had jumped from 4.4 percent to 14.7 percent from March to April. While it fell back to 10.2 percent by July, it remained above the 10 percent peak during the Great Recession in 2007–2009. The S&P 500 remained below the peak on February 19, and home prices were flat. Clearly, 2020 home buyers were justifiably apprehensive about what havoc the coronavirus might heap on them, their friends and family, the value of their homes, their investments, and the country as a whole.

When the survey was mailed out in July 2021, the country had endured over a yearlong battle with COVID-19. While some home buyers' fears from the year before had come to pass, most had not. The outlook was relatively rosy. Coronavirus cases had receded to the lowest level since the start of the pandemic, vaccinations were widely available, and the Delta variant had yet to emerge. At the end of the second quarter 2021, the S&P CoreLogic Case-Shiller U.S. National Home Price NSA Index was up nearly 20 percent from a year earlier and 16.9 percent over the quarter. The S&P 500 had soared nearly 40 percent over the year and 7.5 percent—300 points—in the second quarter. At 5.4 percent, the unemployment rate was nearly half its level the previous year. While many schools and businesses remained physically closed, working from home had become widespread in some sectors with the aid of Zoom. The country was adapting to a new normal.

Figure 9. Survey Questions 33–35 "I Think People Are Becoming More Favorable to . . ."



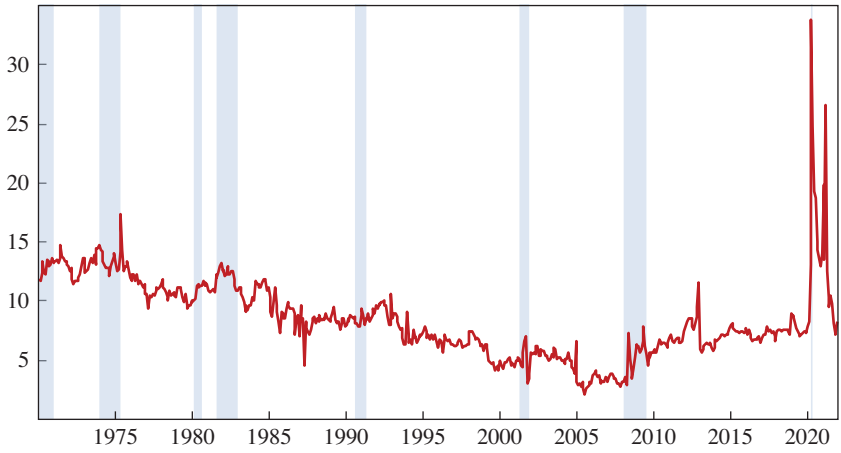
Source: Authors' calculations.

Changes in home buyer's perceptions and expectations between 2020 and 2021 were justifiably stark, the major exception being their view on locational preference. In 2015 we asked whether people were becoming more or less favorable to buying a house in the suburbs, a condo downtown, or renting. We added these questions as the downtown market appeared to be growing in popularity—many believed that retirees would downsize and relocate to the city. While this appeared to be the case through 2019, it changed dramatically with the onset of COVID-19.

Figure 9 shows a slight uptick in the favorability of purchasing a home in the suburbs from 2015 to 2017, followed by a downturn over the following two years. In 2020, preference for buying suburban homes rose nearly 25 percent, while purchasing a downtown condo fell over 25 percent. This trend continued in 2021. Respondents perceived 70 percent of people were amicable to purchasing a suburban home, just 22 percent to buying a condo downtown, and 24 percent to renting.

We supplemented our 2021 survey with additional questions regarding the pandemic, one of which was: "Why do you think home prices have risen so much despite the coronavirus?" While many of the replies mirrored those discussed above, there were additional insights into what drove prices up. The most frequent response to this question was that there was not much impact from COVID-19 on high-wage, white-collar, and tech jobs. Other common threads were that people wanted extra space or home offices because they had been living in cramped quarters during quarantine.

Figure 10. Personal Savings Rate



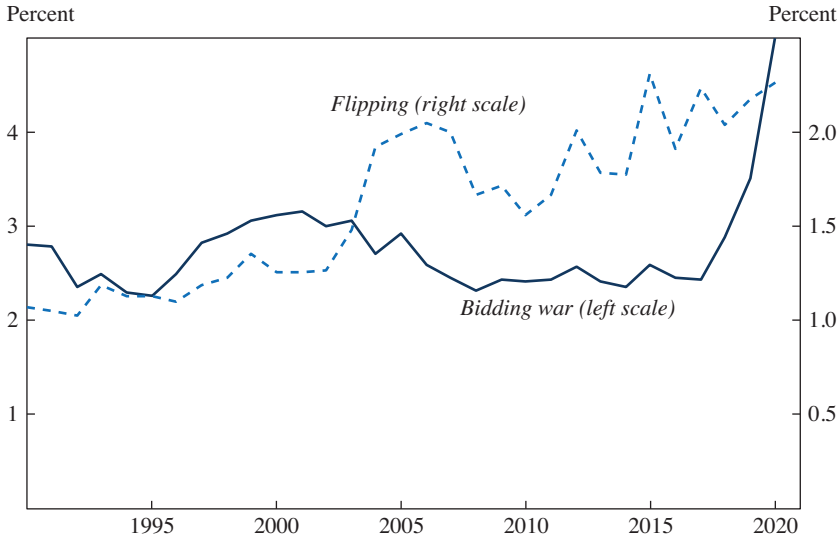
Source: US Bureau of Economic Analysis.
Note: Shaded areas indicate US recessions. The personal savings rate is a percentage of disposable personal income, calculated by the BEA.

Over 87 percent replied that it had become more important to have "a home with one or more office/work rooms." The desire to relocate to the suburbs and change from renting to owning a home were both frequently mentioned. In fact, a record 60 percent were first-time home buyers. In 2020, 28 percent of those replying were age 35–44. This age cohort jumped to a record 35 percent in 2021, over 50 percent of whom reported income of $225,000 or more.

Many noted their spending had slowed during lockdown and they had benefited from stimulus dollars and thus were able to save money for a down payment. Figure 10 shows growth in savings was very atypical during 2020. The personal savings rate, the percentage of disposable income that people save, averaged 16.3 percent in 2020, nearly double the average over the previous fifty years. Therefore, when working from home became the norm, more thought they could be on a path to afford to purchase their first home or upgrade.

When asked why sellers often get multiple offers above the asking price on the day the homes are listed, a record 72.4 percent attributed this to panic buying that caused prices to become irrelevant. These home buyers were getting caught in bidding wars. During their search for a home, a record 50 percent had offered more than the asking price, and 52 percent settled on a price above the asking price. In 2018, we began asking if

Figure 11. ProQuest Web Search for Bidding War and House Flipping, 1991–2021



Source: ProQuest.

Note: ProQuest (www.proquest.com), using search terms home* or hous* and real estate with bid* war* or flip*.

buyers had their offers rejected because someone offered more. In 2021, just 46 percent had their first offer accepted, the remainder had placed bids on other properties that were rejected, and 16 percent had four or more offers rejected before buying their home.

The general vibe in the market had transitioned sharply over the year 2020–2021. The real estate market was hot in 2021, and buyers were well aware of it, with 84 percent describing home prices in the area as rising rapidly. Still, their projected one-year increase in value was just 6.3 percent. In 2021, prices rose 17 percent nationally, 9.4 percent in the second half of the year alone.

VII. The Medium-Term Growth in the Perception of Houses as Speculative Investments

There was a real exuberance in home price expectations in the years leading up to the financial crisis of 2007–2009. This is confirmed by the ProQuest News & Newspapers search shown in figure 11. A "house flipping" narrative took hold, with many stories of fortunes being made by amateur buyers of houses who resell in a matter of months to win great profits. We see that

attention to flipping took hold in 2004 alongside the boom and crested in early 2007, following the 2006 peak in the US housing market. Narratives from before the crisis made some flippers into minor celebrities.

In the run-up to the home price market peak in 2006, there were some viral economic narratives. For example, the reality television show *Flip This House* (2005–2009) on A&E made heroes out of people who got rich buying, fixing up, and quickly reselling houses. One of these was Armando Montelongo, a real estate speculator and motivational speaker, who published a book, *Flip and Grow Rich: The Heart and Mind of Real Estate Investing* (Montelongo 2008).[3] Montelongo was depicted on television as decisive, tough, manly, a fighter, but at the same time down-to-earth. Many viewers of the TV show could identify with him. This show appeared at almost the same time as *The Apprentice* (and spin-offs *The Celebrity Apprentice* and *The New Celebrity Apprentice*) starring Donald Trump (NBC, 2004–2017), which had a similar theme idolizing property speculators.

The attention to house flipping faded away for a few years during the 2007–2009 financial crisis. But a social trend toward "property voyeurism" was not so easily stopped (Harwell 2016). The US housing market began to rise again in 2012, and a number of new reality television shows sprouted up during the following years. Among these were *Property Wars*, which debuted on the Discovery Channel in 2012, and *Flip or Flop* (2013) and *Masters of Flip* (2015) on HGTV. During this flipping revival a number of shows focusing on metropolitan areas were also introduced. Our data show that expectations for future home price increases also rose in 2012, following the same feedback response to actual price increases that we observed in Case, Shiller, and Thompson (2012), as shown for short-term expectations in table 7.

Celebrities like Montelongo were still trying to cash in on investor excitement that was so strong in the years before the 2007–2009 financial crisis. In the 2012 edition of *Flip and Grow Rich* Montelongo wrote, "Housing prices are hitting all-time lows, and this will be the greatest time to make money for the next 40 years. Imagine making more money in 24 months than in the next four decades" (1). He was factually wrong in saying that housing prices were hitting historic lows, but he was right about the real estate market of the coming few years and on how to entice people into speculating in real estate.

3. The title was a parody of Napoleon Hill's 1937 classic self-help book, *Think and Grow Rich.*

These stories, and others like them, have been very much on the minds of the general public, in contrast to the rarer references to institutional investors. People in the television show production business might say that narratives about professionals may not be "aspirational," meaning that they do not feed most viewers' imaginations on how they themselves could really achieve on a high level and ultimately win more respect.

In the 2020s there has been a lot of talk in the news media about institutional investors who are massively investing in homes they will rent out. But these narratives may not have the same currency as those of the flippers. They do not stir emotions among people who have never tried to imagine themselves as institutional investors.

Few home buyers flip houses; instead, they are buying a place to live in. But in making the decision to offer a large sum of money and tying themselves into a long stream of future mortgage payments, they can imagine themselves as being like those speculators.

Most Americans have very little experience with bidding wars. Most of the purchases people make are retail, at a price which is nonnegotiable, ever since the nineteenth century when department stores like Le Bon Marché and Wanamaker's began to advertise that they demand exactly the asking prices to all customers; prices were strictly nonnegotiable.

The idea of bidding wars in real estate, where multiple interested parties make offers on the same property, rose during the late 1990s and early 2000s before tapering off toward the end of the housing boom. Today, bidding wars are a prevalent phenomenon in real estate.

An average of just 2.5 percent of housing and real estate articles mentioned bidding wars between 1991 and 2019. In 2020 it rose to 3.5 percent and surged to 5.0 percent in 2021.

Prior to listing a property, sellers are guided by real estate agents and, increasingly, by internet searches on how to price their home.[4] In a tight market, this often includes strategies to induce a bidding war.[5] Many home buyers, particularly those new to homeownership, face a steep learning curve. Bidding wars are an unfamiliar experience for many at the beginning of their home search but frequently a familiar one by the time they close on a home.

As with sellers, buyers often turn to realtors and the internet for home buying tips. The media is flooded with articles on the best strategies to use

4. See, for example, Zillow, "How to Price Your Home to Sell," https://www.zillow.com/sellers-guide/how-to-price-home-to-sell/.

5. See, for example, Myers (2021).

to increase a buyer's chance of winning a bidding war.[6] You might even think that asking prices should be irrelevant, since actual sale prices occur both below and above the asking price. But they are not irrelevant, since a substantial fraction of sales are exactly at the asking price (Han and Strange 2016). There is an inscrutable psychological game developing in the housing market that may increasingly favor speculative impulses (Gadd 2006).

Redfin, a full-service real estate brokerage website, reports bidding war rates based on offer competition. Among Redfin realtors, the rate has increased dramatically since they began compiling this information. In April 2020, near the start of the COVID-19 crisis, a seasonally adjusted 33 percent of Redfin offers faced competition (Katz 2022). By January 2021 this rate had risen to 61 percent, and in January 2022 it reached 70 percent. One can't help but wonder whether growth in bidding wars is a contributing cause to the rise in home prices during the pandemic.

VIII. Conclusion

Our analysis of our surveys of home buyers in 1988 and from 2003 to 2021 shows that home buyers' expectations are fairly rooted in reality for the short run, underreacting in the short run, but given to flights of fantasy for the longer run. The shorter-run expectations were pretty much on target throughout the period. This is not really a surprise: looking at plots of the data, like those in figure 1, we see that home prices are quite smooth through time and hence easily forecastable by simple extrapolation for a short time after the survey, in sharp contrast to stock prices, which somewhat resemble random walks. But forecasting the longer run presents a real challenge and tends sometimes to go to extremes that are at odds with reality.

Since the strong uptrend in home prices that started in 2012 and strengthened with the COVID-19 pandemic in 2020 is not associated with high ten-year expectations for price increase, and since home buyers mostly stay in their homes for years or decades, we would not call the experience a bubble, at least not in the classic sense.

But it resembles a bubble in the sense that it is driven by a kind of excitement or fear of missing out (FOMO, in today's internet lingo). The excitement is associated with having to deal with bidding wars and worries about being outbid if one does not bid aggressively enough. The public mood among those actively bidding is one of fear of being jilted, losing a house they may have fallen in love with to a more aggressive competitor.

6. See, for example, LaPonsie (2021).

Forecasting house prices at this point in history is not just a matter of judging the progress of a hypothetical bubble. While the rapid increase in home prices is a cause for concern, forecasters must go beyond simple models to forecasting the COVID-19 epidemic and its future variants, or forecasting international tensions such as those raised by the Russian invasion of Ukraine in February 2022 and Vladimir Putin's veiled threats to use atomic weapons against nations who support Ukraine. They must also consider the change in supply of housing and in communications technology, the changing geographical distribution of business activity, and the evolution of popular narratives about these things.

In concluding, we should also remind that the patterns of expectations we describe are not immutable. The volatility in these popular expectations can be made less impactful if financial institutions are improved. The home price futures market that we advocated (Case, Shiller, and Weiss 1993), which was created by the Chicago Mercantile Exchange in 2006 based on the S&P CoreLogic Case-Shiller home price indexes, does allow an investor to hedge home price risk, though to this day the market is not very liquid. A number of attempts have been made over the last thirty years to make possible better risk management of home price volatility (Fabozzi, Shiller, and Tunaru 2020). Eventually, such risk management may make homeownership less of a gamble in the future.

ACKNOWLEDGMENTS The authors are indebted to the US National Science Foundation, which funded the 1988 survey, and to the Yale School of Management, with a grant from Whitebox Advisors which supports our recent surveys. Cathy Adrado, Laurence Black, Daniel Boston, Leigh Clark, Zachary Dewill, Milad Nozari, Virginia Shiller, David Still, and Olga Vidisheva provided help for this research.

This paper is dedicated to the memory of Karl "Chip" Case, our coauthor on the paper the three of us published ten years ago in *Brookings Papers on Economic Activity*, which was a predecessor. Chip contributed heavily to our joint work before his death on July 15, 2016.

References

Board of Governors of the Federal Reserve. 2020. "Federal Reserve Issues FOMC Statement." Press release, March 23. https://www.federalreserve.gov/newsevents/pressreleases/monetary20200323a.htm.

Brunnermeier, Markus, and Christian Julliard. 2008. "Money Illusion and Housing Frenzies." *Review of Financial Studies* 21, no. 1: 135–80.

Case, Karl E., and Robert J. Shiller. 1988. "The Behavior of Home Buyers in Boom and Postboom Markets." *New England Economic Review* (November/December): 29–46.

Case, Karl E., Robert J. Shiller, and Anne K. Thompson. 2012. "What Have They Been Thinking? Homebuyer Behavior in Hot and Cold Markets." *Brookings Papers on Economic Activity*, Fall, 265–98.

Case, Karl E., Robert J. Shiller, and Allan N. Weiss. 1993. "Index-Based Futures and Options Markets in Real Estate." *Journal of Portfolio Management* 19, no. 2: 83–92.

Demyanyk, Yuliya, and Otto Van Hemert. 2011. "Understanding the Subprime Mortgage Crisis." *Review of Financial Studies* 24, no. 6: 1848–80.

Draghi, Mario. 2012. Speech given at the Global Investment Conference, London, July 26. https://www.ecb.europa.eu/press/key/date/2012/html/sp120726.en.html.

Fabozzi, Frank J., Robert J. Shiller, and Radu Tunaru. 2020. "A 30-Year Perspective on Property Derivatives: What Can Be Done to Tame Property Price Risk?" *Journal of Economic Perspectives* 34, no. 4: 121–45. https://www.aeaweb.org/articles?id=10.1257/jep.34.4.121.

Ferreira, Fernando, and Joseph Gyourko. 2011. "Anatomy of the Beginning of the Housing Boom: U.S. Neighborhoods and Metropolitan Areas, 1993–2009." Working Paper 17374. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w17374.

Gadd, Jane. 2006. "Picking a Price Can Be Perilous." Globe and Mail, June 23. https://www.theglobeandmail.com/real-estate/picking-a-price-can-be-perilous/article711489/.

Gorton, Gary B. 2010. *Slapped by the Invisible Hand: The Panic of 2007*. Oxford: Oxford University Press.

Han, Lu, and William C. Strange. 2016. "What Is the Role of the Asking Price for a House?" *Journal of Urban Economics* 93:115–30.

Harwell, Drew. 2016. "How 'House Hunters' Became the Most Unstoppable Juggernaut on TV." *Washington Post*, January 25. https://www.washingtonpost.com/news/business/wp/2016/01/25/how-house-hunters-became-the-most-unstoppable-juggernaut-on-tv/.

Katz, Lily. 2022. "January Was the Most Competitive Month on Record for Homebuyers." Redfin News, February 23. https://www.redfin.com/news/real-estate-bidding-wars-january-2022/.

LaPonsie, Maryalene. 2021. "How to Win a Bidding War on a House." *U.S. News and World Report*, June 7. https://realestate.usnews.com/real-estate/articles/how-to-win-a-bidding-war-on-a-house.

Mathis, Jérôme, James McAndrews, and Jean-Charles Rochet. 2009. "Rating the Raters: Are Reputation Concerns Powerful Enough to Discipline Rating Agencies?" *Journal of Monetary Economics* 56, no. 5: 657–74.

Mian, Atif, and Amir Sufi. 2009. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *Quarterly Journal of Economics* 124, no. 4: 1449–96.

Montelongo, Armando. 2008. *Flip and Grow Rich: The Heart and Mind of Real Estate Investing*. Armando Montelongo Worldwide; 3rd edition, 2012.

Myers, Emily. 2021. "Pulling Off a Bidding War: A Seller's 8-Step Guide." Blog post, Brick Underground, June 9. https://www.brickunderground.com/blog/2014/05/managing_bidding_war.

Comments and Discussion

COMMENT BY

ADAM M. GUREN It is unusual to be asked to discuss a paper that is already a classic, but that is the predicament in which I find myself. This paper is a useful update on the pathbreaking and influential work in the authors' prior paper, Case, Shiller, and Thompson (2012). The natural place to start discussing this ten-year retrospective is thus with a ten-year retrospective discussion of the original paper. Explaining why it has been so influential and what its impact means for the current paper will give me an opportunity to discuss its methodology and the related literature and provide a brief user's guide to their data. I will then turn to discussing the analysis of the last ten years in housing markets and particularly the current pandemic housing market.

My overall message is that the authors should be applauded for their important contribution: the field of housing economics is unquestionably better due to their adding survey expectations to our tool kit. I also think their big idea—that high, long-run expectations can be used to diagnose a housing bubble, much like a yield curve inversion is used to predict a recession—is a useful one, although I think that survey evidence on expectations should be used in conjunction with other evidence rather than on its own.

SURVEY EVIDENCE ON HOUSE PRICE EXPECTATIONS

The influence of Case, Shiller, and Thompson (2012). When the authors of this paper began surveying home buyers on their expectations about the future path of house prices in the late 1980s, the idea that one would ask economic agents about their behavior and expectations was outlandish. Even in 2003, when they revived the survey, it was novel. Today, however, survey evidence on expectations is widespread and accepted as a crucial tool.

Brookings Papers on Economic Activity, Spring 2022: 345–366

The authors deserve a great deal of credit for pioneering and legitimizing survey evidence on expectations in housing markets. Their survey was, to my knowledge, the first to go beyond the Michigan Surveys of Consumers' question on whether it is a good or bad time to buy a house and actually ask for expectations of house price growth over various horizons as well as buyers' subjective views about the state of the market. Their work helped demonstrate the value of survey evidence for understanding housing markets and cycles and make its use commonplace, although survey evidence must be taken with a grain of salt and evaluated carefully.[1]

The original paper helped launch a large body of literature, which has been recently and comprehensively surveyed by Kuchler, Piazzesi, and Stroebel (2022). Given space constraints, I only note a few highlights relevant to the current paper here.

First, several key observations that Case, Shiller, and Thompson (2012) made in their initial paper, which included only ten years of data for four metro areas, have been shown to be key features of expectations after more comprehensive analysis. In particular, Case, Shiller, and Thompson (2012) hypothesized that there was the underreaction of short-term (one year ahead) expectations and overreaction of longer-run (ten years ahead) expectations, a finding that shows up again in section V of this update. Recently, Armona, Fuster, and Zafar (2019) used a novel informational experiment to show convincingly and causally that short-run expectations underpredict the degree of short-run momentum and long-run expectations do not fully account for mean reversion in house prices. Similarly, Case, Shiller, and Thompson (2012) postulated that sentiment about house prices spreads through the media and by word of mouth. Bailey and others (2018) validated this using Facebook data, showing that individuals' expectations are formed in part by the price appreciation of their out-of-town friends. Overall, with ten years of hindsight, the original paper seems uncannily prescient.

Second, since Case, Shiller, and Thompson (2012), surveys like theirs have proliferated both in the United States and abroad, which is important for two reasons. First, having more surveys and additional countries provides both more data to do the type of analysis for which the authors advocate and a richer baseline set of facts. Second, in the United States there

1. Influential work by Coibion and Gorodnichenko (2012) that uses survey evidence on inflation expectations to discriminate between models also played an important role in popularizing the use of survey evidence outside of housing markets. The wider acceptance of survey evidence in macroeconomics played a role in its growing acceptance in housing economics.

are now high-quality surveys of house price expectations by the Michigan Surveys of Consumers and the Federal Reserve Bank of New York which come out monthly and with very little lag. These surveys make data on house price expectations a viable real-time tool for policymakers and economic forecasters. Case, Shiller, and Thompson (2012) famously pointed out that in the 2000s boom, long-run expectations ballooned to the point that they were higher than short-run expectations, which they said indicated a bubble. The presence of these sorts of data will hopefully aid in identifying bubbles as they occur.

Finally, Case, Shiller, and Thompson (2012) helped motivate a literature that uses nonstandard expectations to explain the 2000s cycle. At this point, essentially every legitimate explanation of the 2000s housing cycle ascribes a significant role to overoptimistic or out-of-line expectations, although there remains disagreement on the relative role of out-of-line expectations relative to other explanations like a credit supply expansion and on the source of the out-of-line expectations. Furthermore, many papers use the Case-Shiller-Thompson (CST) data to help discipline explanations and models of what happened in the boom and bust and to discriminate between various models of nonrational beliefs, a point to which I return below.[2]

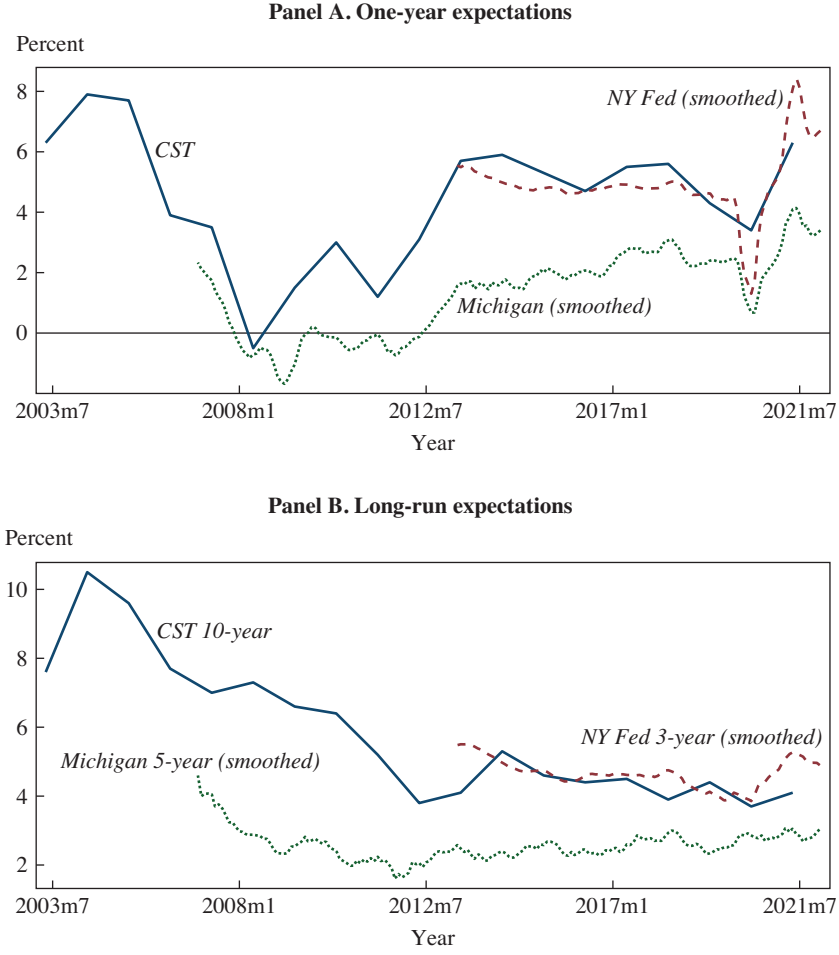
What is unique about their survey? The success of the original paper in inspiring several similar surveys somewhat limits the novelty of the findings about the last ten years in this update. While the authors point out that their survey is unique and preferable to others because it has the longest panel and because it covers recent active market participants rather than the public at large, the surveys from the Federal Reserve Bank of New York and Michigan Surveys of Consumers come out more frequently, are weighted to be representative of a full population, and use modern best practices in survey design.

Figure 1 compares these three data sources. Panel A shows one-year expectations and panel B shows long-run expectations. In both panels, the thick solid lines show the CST data in this paper (I add the late and great Chip Case's name to acknowledge his contribution), the dashed lines show data from the Federal Reserve Bank of New York Survey of Consumer Expectations (NY Fed), and the dotted lines show data from the Michigan Surveys of Consumers (Michigan).

These three surveys have different survey methodologies, phrase the questions differently, ask about various time horizons, and survey different

2. See, for example, Burnside, Eichenbaum, and Rebelo (2016), Kaplan, Mitman, and Violante (2020), and Chodorow-Reich, Guren, and McQuade (2021).

Figure 1. Comparison of House Price Expectation Surveys



Sources: Federal Reserve Bank of New York Survey of Consumer Expectations and University of Michigan Surveys of Consumers.

Note: The Case-Shiller-Thompson (CST) data are from table 3 of the paper, and the longer-run expectation is average annual house price growth over the next ten years from the survey date. The Federal Reserve Bank of New York (NY Fed) data are monthly data smoothed using a five-year moving average, and the longer-run expectation is price growth between twenty-four and thirty-six months from the survey date. The Michigan survey data are monthly data smoothed using a five-year moving average, and the longer-run expectation is average annual price growth over the next five years from the survey date. Data are accurate as of February 2022.

groups. These differences in survey design and phrasing can matter immensely. Specifically, the authors mail surveys to a random selection of recent home buyers in four distinct markets in the spring of each year, asking them for one-year-ahead and ten-year-ahead annual average appreciation beginning in 2003.[3] The NY Fed uses an internet survey of a nationally representative group of household heads and asks them about national house price appreciation in the next year and from twenty-four to thirty-six months from the survey date beginning in 2014.[4] The Michigan survey is a nationally representative telephone survey of households that asks them about the appreciation of "homes like yours in your community" over the next year and annual averages over the next five years beginning in 2007.[5]

Despite the differences, one can see that both the one-year and longer-run expectations are similar for the periods they overlap. In particular, the authors' and NY Fed survey's expectations are close to overlapping, with a more prominent drop early in the pandemic and spike late in the pandemic for the higher-frequency NY Fed data. The Michigan data, by contrast, generally give lower average expectations but similar time paths. Nonetheless, the key patterns that the authors highlight in this paper, namely, the fact that longer-run and shorter-run expectations largely overlap since 2012 and do not appear out of line in the pandemic, are both visible in the NY Fed and Michigan surveys. The fact that in the bust short-term expectations fall by more than long-term expectations is visible in both the CST and Michigan data.

3. The authors ask, "How much of a change do you expect there to be in the value of your home over the next 12 months?" and "*On average* over the next ten years how much do you expect the value of your property to change *each year*?" (their emphasis, which was added starting in the 2012 survey). They have a response rate of 12–44 percent. Their survey asks only about house prices.

4. The NY Fed survey says, "Think about home prices nationwide" and asks, "*Over the next 12 months*, by about what percent do you expect the average home price to increase/decrease?" and "*Over the 12-month period between* [twenty-four months from survey date] *and* [thirty-six months from survey date], by what percent do you expect the average home price to increase/decrease?" (their emphasis). The survey is weighted to be representative given response rates. The survey asks about a broad range of expectations, with the house price questions coming immediately after questions about income and credit availability in the middle of the survey.

5. The Michigan Surveys asks, "By about what percent do you expect prices of homes like yours in your community to go up/down, on average over the next twelve months?" and "By about what percent per year do you expect prices of homes like yours in your community to go up/down, on average, over the next five years or so?" The survey is weighted to be representative given response rates. The survey asks about a broad range of expectations, with the house price questions coming after questions about inflation, and specifically gas prices, toward the end of the survey.

What makes the CST data unique is that it is the only survey that covers the entirety of the 2000s boom and bust. This cycle is the largest and most consequential on record, so having data that show just how out of line expectations—and in particular longer-run expectations—were in the boom is crucial to being able to use expectations data to diagnose a bubble in real time going forward. Indeed, the authors' observation that long-run expectations do not seem out of line in the pandemic boom is only revealing in comparison to their findings on the 2000s. The NY Fed survey may capture the later period, but since we do not know what this time series would look like in a significant boom and bust, it is hard to know how to interpret the COVID-19-era data. The same goes to a lesser extent for the Michigan data, which do not cover the 2000s boom.

That being said, given that data are released in near real time, the NY Fed and Michigan surveys are the early warning system for policymakers and economic forecasters. To maximize the impact of their research and big ideas about how survey expectations of house prices can be used to assess the direction of the housing market, I hope that Shiller and Thompson can work with the NY Fed and Michigan to compare survey designs and questions. By asking each other's questions with each other's phrasings for several years going forward—and possibly asking lab participants to answer multiple different survey questions and phrasings in multiple different scenarios—one can get to the bottom of whether these surveys behave differently due to the groups surveyed, the phrasing and sequencing of the questions, or other factors. This will help us ascertain how the real-time surveys might look in a 2000s-like housing cycle and help policymakers assess the trajectory of the housing market with this sort of survey data. In other words, by treating these other surveys as complements, not competitors, I think Shiller and Thompson can dramatically increase the influence and use of the type of survey data they pioneered.

A user's guide to the CST data. There are two important things that users of the CST data should know.

First, the CST data suffer particularly in the boom (2003–2005) period from extremely high reported ten-year expectations. While the authors argue that 10 percent expected price inflation over the next ten years is not out of line with what actually happened over the prior ten years, any model that attempts to match the level of expected ten-year appreciation they find will dramatically overpredict the size of the boom. Indeed, all papers that use the CST data as a calibration target that I know of find a way to artfully dodge the ten-year house price expectations in 2004 and 2005 for this

very reason—either by using the one-year expectations, by using an average of many years of the ten-year expectations, or by using the CST ten-year expectations starting in 2006.

My concern, and the concern of David Laibson (2012), who focused his discussion of the original Case, Shiller, and Thompson paper on this issue, is that some of this is due to respondents misunderstanding the question. Prior to 2012, the survey did not underline and bold that the ten-year expectation was supposed to be "on average" and a growth rate for "each year," and Laibson argues that some households misread this question in particular by conflating the average and total return. Given this, Laibson writes that the ten-year expectations "cannot be interpreted literally" (301). The authors clearly took this seriously, as they explored this in the 2013 survey and found that 22 percent misunderstood the question. I share some of Laibson's hesitance, but I am pleased to report that in this ten-year retrospective paper the authors do a better job adjusting for the sorts of survey confusion that concerned Laibson prior to 2012. Rather than reporting raw 10 percent trimmed means (dropping the highest and lowest 5 percent of responses and then calculating a mean), the authors now replace cases where the respondent gave a ten-year annual average expectation more than ten times their one-year expectation with the one-year expected values and then calculate a 10 percent trimmed mean. This brings down some of the more extreme expectations—for instance, in 2004 and 2005 in Orange County, the average expectation for annual appreciation over the next ten years is 13.3 percent and 10.4 percent, respectively, rather than 17.4 percent and 15.2 percent. While these results should still be taken with a grain of salt and used carefully, the new figures are preferable.

Second, given the improvements in calculating the ten-year expectations in the new version, I urge researchers to use the updated expectations data from this 2022 version rather than the 2012 paper. That being said, I think there are still potential improvements. For instance, it is not clear that replacing the ten-year expectation with the one-year expectation when the ten-year expectation is implausible is desirable. For this reason, I hope that Shiller and Thompson are able to release anonymized micro data so that researchers can implement their own trimming procedures as appropriate (not to mention analyze things like disagreement that one can only consider with micro data).

THE 2012–2020 REBOUND Part of the authors' analysis focuses on the "second" or "current" boom from 2012–2020. They compare this boom—in which short-run and long-run price expectations have remained stable

and in line with mortgage interest rates—to the 2000s boom when short- and particularly long-run expectations skyrocketed. One interpretation of their findings is that expectations were more rational in the second boom.

I want to present an alternate view, which I develop with Gabriel Chodorow-Reich and Tim McQuade in a recent paper (Chodorow-Reich, Guren, and McQuade 2021), that the 2012–2020 boom is not a second boom but instead the rebound phase of a single, twenty-year boom-bust-rebound cycle. We begin with the observation that in the cross section, areas with the largest booms (1997–2006) and busts (2006–2012) also had the largest rebounds (2012–2019). Indeed, the bust and the rebound are as highly correlated in the cross section as the bust and boom are. Furthermore, the boom is highly correlated with the overall 1997–2017 boom-bust-rebound price growth, with an R^2 of 0.62. The extremely high correlations across the three phases are indicative of a single boom-bust-rebound cycle rather than a boom-bust followed by a second unrelated boom. Furthermore, high correlation between the boom and longer-term price growth from 1997 to 2019 is suggestive of the boom being an overreaction to real improvements in fundamentals, an idea we explore systematically both in the data and using a model in the remainder of our paper.

We first pursue this interpretation of fundamental improvements driving long-run price growth empirically using a structural urban framework. We extract a city-level fundamental as a function of instruments for income, amenities, and supply and show that our estimated fundamental is correlated not only with long-run house price growth but also with the amplitude of the boom-bust-rebound cycle and the severity of the foreclosures crisis in the bust.

We then write down a model of a fundamentally rooted house price cycle. In the model, a single improvement in the drift term of the dividend to living in a city in the late 1990s leads to a boom-bust-rebound pattern consistent with the data. Intuitively, the boom is generated by overoptimism about the fundamental improvement, the bust occurs as beliefs correct, bringing down prices and leading to price overshooting due to foreclosures. Finally, the rebound occurs as foreclosures recede and prices converge to a new, higher-growth, balanced growth path. In the model, overoptimism occurs due to diagnostic expectations, which are nonrational expectations developed by Bordalo and others (2019) that embed a tractable formalization of Kahneman and Tversky's representativeness heuristic. The representativeness heuristic is that people tend to overweight the likelihood of a trait in a class when that trait has a higher likelihood in a class than in a reference population; for instance, people tend to overestimate the share of Irish with

red hair because red hair is more prevalent among the Irish. In the context of asset prices, the reference population is the full history of observed dividends and the class is recently observed dividends, with inference over the dividend drift rate. As people observe higher dividends, they overweight the probability of a very high dividend growth state, leading their long-run house price expectations to rise significantly. As people get more and more data, they realize their error, and their beliefs converge to the rational belief gradually from above. Combined with the overshooting on the downside from foreclosures, this delivers a boom-bust-rebound, which we show is quantitatively consistent with the cross section of boom-bust-rebounds across groups of cities. I see nothing in the authors' analysis of the 2012–2020 boom that is inconsistent with this story and consider it useful to think of the 2012–2020 boom this way in interpreting the authors' results.

Chodorow-Reich, Guren, and McQuade (2021) is also a good example of how the literature has used the CST expectation survey data to discriminate between various models of nonrational beliefs and discipline macro models of the 2000s housing cycle. Indeed, one of the main reasons we use diagnostic expectations is because they are consistent with the CST observation that long-run expectations do not overshoot in the bust and instead converge smoothly from above; most other candidate models of expectations do not give this prediction.[6] The findings in here thus not only help reject rationality but also help narrow down the set of plausible nonrational models.

THE PANDEMIC HOUSING MARKET Much of the analysis in the paper concerns the recent surge in house prices seen since the onset of the COVID-19 pandemic. As mentioned above, the authors suggest that high long-run expectations can be used to diagnose a housing bubble in real time much like a yield curve inversion is used to predict a recession. They argue that ten-year expectations have not grown rapidly and so "we would not call the experience a bubble, at least not in the classic sense." That being said, they do hedge themselves a bit by saying the market "resembles a bubble in the sense that it is driven by a kind of excitement or fear of missing out."

At the risk of making an inaccurate prediction, I agree with the authors: the pandemic price surge does not look like the bubble we experienced in the 2000s, so it is unlikely we will experience a correction like the one we experienced in the 2000s. In coming to this conclusion, the authors' observation that long-run expectations have not increased significantly is an

6. Diagnostic expectations also create an independence between the amplitude of the cycle and the length of the boom and bust, which we observe in the cross-city data.

important data point, but it is not the only one. It is also important to note the lack of rapid credit expansion or speculation by short-term traders or house "flippers." Various measures such as loan-to-value ratios, payment-to-income ratios, and credit scores of new mortgages suggest that we have not seen the type of credit expansion that occurred across the spectrum of borrower quality in the 2000s. For instance, one summary measure, the Mortgage Credit Availability Index published by the Mortgage Bankers Association, saw a huge expansion and contraction in the 2000s but has barely budged in the pandemic. This is important because authors like Greenwood and others (2022) have argued that rapid expansions in credit together with asset prices are predictive of bubbles and financial crises. Similarly, we do not seem to see a surge in the number of non-owner-occupant purchasers that we saw in the 2000s, which is a real-time proxy for the presence of speculators.[7] This sort of speculative activity is another hallmark of asset bubbles. Overall, I agree with the authors that policymakers and economic forecasters should use measures of long-term house price expectations in diagnosing a bubble, but I think they should be used in conjunction with other indicators rather than in a vacuum.

The other reason I am skeptical that the pandemic housing market is an expectation-fueled bubble is that there are good reasons to think that the pandemic has increased housing demand and constrained housing supply. On the demand side, the pandemic increased demand for housing space. Households began to work from home. City dwellers learned the value of additional space, particularly outdoors. Younger families decided to move to the suburbs sooner. At the same time, older households decided to age in place rather than downsizing or moving to senior living or a nursing home. All of these trends dramatically increased the demand for housing space. On the supply side, several factors have conspired to limit supply. For existing homes, the lack of downsizing by older households limited the supply of existing homes and led to record-low inventories of existing for-sale homes. Construction was also limited by material availability and supply chain disruptions as well as strong labor markets that drove up wages for construction workers, and particularly for skilled construction workers, who are in short supply. Immigration restrictions have also helped to drive up construction wages. The pandemic-induced expansion in demand and inelastic supply together led to a surge in house prices.

7. "Speculators" do not include institutional investors who are buying properties to hold and rent out. The market share of this type of investor has grown recently.

Of course, the fact that we will likely not experience a correction like the one we experienced last time does not mean there will not be a different type of correction. We tend to fight the last war but there are other factors to think about in the pandemic housing market that may not be as easily diagnosed using survey expectations. In particular, in assessing the risks of a different correction, housing economists should be paying attention to future supply and preferences for housing space.[8]

First, an eventual supply response could lead to a correction. The housing market currently appears to be supply constrained even in cities typically considered to be long-run elastic. Eventually supply should respond and construction should accelerate. But when and by how much? And perhaps more importantly, are market participants accounting for a medium-run supply response when forming their expectations today? Or are market participants neglecting the future supply response (Greenwood and Hanson 2015)? Joseph Gyourko is an expert on housing supply and discusses these questions at length in the next comment, so I will be brief: I share his concerns about how much we can learn from survey expectations when those expectations may not incorporate future supply responses fully.

Second, it is unclear whether the pandemic-induced change in preferences will reverse or be permanent. As I discussed above, the pandemic led to a significant increase in demand for housing space due to everything from working from home to people spending more time at home and valuing having more space to the desire to have outside space to older households wanting to age in place. One could imagine a world in which this reverses and prices decline significantly. One could just as easily imagine a world in which many of these changes in tastes are long-lived and housing demand stays strong.

In thinking about the pandemic, I am often drawn to the work of Malmendier and Nagel, who show that living through a traumatic economic event such as the Great Depression (Malmendier and Nagel 2011) or Great Inflation (Malmendier and Nagel 2016) has long-lasting impacts on economic agents' preferences, risk tolerances, and expectations. I suspect the COVID-19 pandemic will have similar long-run effects on preferences, but it is hard to know exactly how and whether preferences relevant to housing

8. In the very short run, inflation and interest rates are also a risk for the housing market. At the time of writing, mortgage rates have surged from about 2.8 percent in the fall to 5.3 percent in May 2022. Rising mortgage rates will put a damper on housing demand and could cool the market off and potentially lead to a correction.

markets like working from home and wanting more space will continue long-term. Economists are generally bad at forecasting changes in preferences. I thus think we need to be humble in our forecasts of housing prices and housing demand more generally.

I want to conclude where I began: the authors should be applauded for introducing and popularizing survey expectations about house prices as an important tool in analyzing housing markets. The more data on this subject that are available the better. I hope that going forward the authors are able to open up their tremendous treasure trove of data—including micro data—to researchers so that we can better understand house price expectations and their role in shaping the dynamics of housing markets.

REFERENCES FOR THE GUREN COMMENT

Armona, Luis, Andreas Fuster, and Basit Zafar. 2019. "Home Price Expectations and Behaviour: Evidence from a Randomized Information Experiment." *Review of Economic Studies* 86, no. 4: 1371–410.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018. "Social Connectedness: Measurement, Determinants, and Effects." *Journal of Economic Perspectives* 32, no. 3: 259–80.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. 2019. "Diagnostic Expectations and Stock Returns." *Journal of Finance* 74, no. 6: 2839–74.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo. 2016. "Understanding Booms and Busts in Housing Markets." *Journal of Political Economy* 124, no. 4: 1088–147.

Case, Karl E., Robert J. Shiller, and Anne K. Thompson. 2012. "What Have They Been Thinking? Homebuyer Behavior in Hot and Cold Markets." *Brookings Papers on Economic Activity*, Fall, 265–98.

Chodorow-Reich, Gabriel, Adam M. Guren, and Timothy J. McQuade. 2021. "The 2000s Housing Cycle with 2020 Hindsight: A Neo-Kindlebergerian View." Working Paper 29140. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29140.

Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy* 120, no. 1: 116–59.

Greenwood, Robin, and Samuel G. Hanson. 2015. "Waves in Ship Prices and Investment." *Quarterly Journal of Economics* 130, no. 1: 55–109.

Greenwood, Robin, Samuel G. Hanson, Andrei Shleifer, and Jakob Ahm Sorensen. 2022. "Predictable Financial Crises." *Journal of Finance* 77, no. 2: 863–921.

Kaplan, Greg, Kurt Mitman, and Giovanni L. Violante. 2020. "The Housing Boom and Bust: Model Meets Evidence." *Journal of Political Economy* 128, no. 9: 3285–345.

Kuchler, Theresa, Monika Piazzesi, and Johannes Stroebel. 2022. "Housing Market Expectations." Working Paper 29909. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29909.

Laibson, David. 2012. Comment on "What Have They Been Thinking? Homebuyer Behavior in Hot and Cold Markets," by Karl E. Case, Robert J. Shiller, and Anne K. Thompson. *Brookings Papers on Economic Activity*, Fall, 299–301.

Malmendier, Ulrike, and Stefan Nagel. 2011. "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?" *Quarterly Journal of Economics* 126, no. 1: 373–416.

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

COMMENT BY

JOSEPH GYOURKO It is a pleasure and honor to comment on this paper. Not only is it of broad interest to professional economists, but it can profitably be read by policymakers and practitioners in housing markets. The ongoing data collection effort that underpins the paper was visionary when begun in 1988 and now constitutes a valuable public good for the economics profession. The repeated cross sections date back far enough in time to cover more than a full housing cycle—spanning the long boom leading up to the global financial crisis, the subsequent great decline, the long recovery, and now the sharp upsurge in prices during the COVID-19 crisis. The length of that time span makes it unique compared to newer surveys of price expectations. I only wish that Chip Case, an original collaborator on this research program, was here to help present this second ten-year review of results for Brookings. He is much missed, especially by those of us who study housing markets.

The paper is well written and provides detailed descriptions of many of its key results. I see no reason to summarize or critique much of that material, although I do provide specific commentary in select instances. Hence, most of my comments are wide-ranging in nature. They include a suggestion to link this paper's survey and results to other research that was initiated more recently and often inspired by the Case, Shiller, and Thompson series. Doing so would help the economics profession see what it can learn from better integrating this effort with that of others. Hopefully, the result will be greater than the simple sum of the individual parts (i.e., of the different surveys of expectations). I also include a detailed discussion of the COVID-19 period. The authors provide substantial detail on this period, too, as it has been a remarkable time for housing markets. I do not think there is much to improve upon in terms of their discussion of the 2020 and

2021 survey results, but I do hope to influence questions they ask in next year's survey. One of the great strengths of this survey is that respondents are not only asked what they think will happen to house prices in the short and long terms. Supplementary questions provide the opportunity to dig deeper into the reasons people provide to justify their price expectations. They hold the potential for us to better understand the working of housing markets beyond what expectations themselves provide.

LINKING TO OTHER SURVEYS AND RESEARCH My first suggestion is more for the broader economics profession than it is for the authors of the paper, that is, to begin linking the results presented and discussed in the paper to other, newer surveys on price expectations. For example, the housing module of the Federal Reserve Bank of New York's Survey of Consumer Expectations was initiated in 2014, so its time series is not nearly as long as that discussed in this paper.[1] However, that survey explicitly tries to achieve a nationally representative sample that includes those who did not purchase a home, so there are advantages to its approach. Comparing and contrasting the two sets of findings and methodologies would be a useful first step. Ultimately, it may prove possible to integrate the data, by which I mean using the common 2014–2021 period across these two surveys to adjust the series reported in Shiller and Thompson backward in time. That is a potentially complex undertaking which is not appropriate for this paper, but I encourage the authors to begin working with the people conducting other surveys to see what we might be able to learn from closer coordination going forward.

A related comment on linking to more recent work that the research program underlying this paper has inspired over time pertains to more recent developments in the analysis and estimation of whether expectations are rational. Table 5 in the paper reports results for short-run expectations. I am not an expert in this specific area, but there are newer methodologies described in Kuchler, Piazzesi and Stroebel (2022), for example. I do not believe that the results reported here will change materially after employing a different specification or econometric technique, but I suspect that the more we can link surveys and analytical approaches, the better.

SHORT-RUN AND LONG-RUN EXPECTATIONS Short-run (one year) and long-run (ten year) house price growth expectations typically are the paper's headline results. I will not reiterate the authors' discussion of them in the paper, but I do think it noteworthy that the nature of their conclusions about

1. Federal Reserve Bank of New York, Center for Microeconomic Data, https://www.newyorkfed.org/microeconomics/sce/housing#/

short-term expectations in particular has not changed with the passage of time. This is not to say the level of the expectations themselves has been constant. Quite the contrary, in fact. Rather, it is meant to imply that the characteristics displayed by one-year-ahead expectations have not materially changed since the last report of these data to Brookings in 2012 (Case, Shiller, and Thompson 2012). This means they remain more volatile in nature, have some backward-looking element to them, but clearly contain information about the future above and beyond what is reflected in other regression controls such as lagged price growth. More simply put, short-term expectations look to be rationally founded, and the long recovery after the global financial crisis and the COVID-19 pandemic have not changed that perspective.

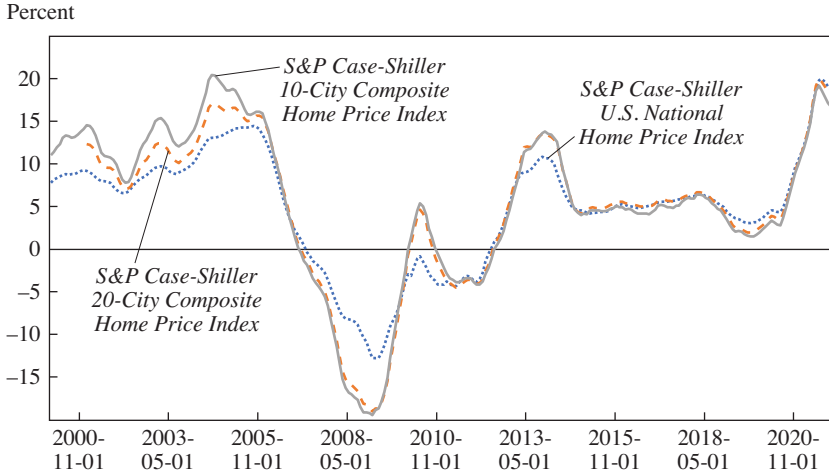
The 2012 paper did not conclude the same about ten-year expectations. Specifically, the authors note that long-term expectations appear too optimistic in the 2003–2006 period, with a large 200–500 basis point gap between those expectations and the thirty-year mortgage rate existing during that period (see figure 5 in the paper for the plot of these two series). That gap was emphasized as the basis for expecting high financial returns to homeownership and thus was a potential explanation for any price bubble around that time. A much more modest gap opened up in 2021, but currently forecasted Federal Reserve interest rate rises seem likely to ensure it will not widen in 2022. Hence, longer-term expectations now look less bubbly to the authors for this reason and because they have not been trending up over recent years. The slow, longer-run movement in long-term expectations is similar across the two halves of the time series. What is striking now is that long-term price growth expectations are markedly below short-term expectations in the most recent survey for 2021.

LEARNING MORE ABOUT HOUSING MARKETS DURING THE COVID-19 CRISIS

I appreciate that the section on the COVID-19 era has been expanded in this final version. It is well worth the added detail. One reason is that price appreciation has been stunning, as indicated by my plot of the Case-Shiller (nominal) repeat sales series (figure 1). Price appreciation very quickly reached growth rates not seen since 2004–2005. Moreover, it is more widespread, as indicated by the near overlap of the national, 10-City Composite and 20-City Composite series since COVID-19 hit the United States. There was much more heterogeneity across markets in the run-up to the global financial crisis.

The data on expectations here are interesting as usual. The results for 2020, which reflect answers from buyers who knew that COVID-19 existed but bought before the virus really hit America, show short-run expectations

Figure 1. Case-Shiller Repeat Sales Index



Source: FRED, retrieved from https://fred.stlouisfed.org/series/. Reproduced with permission from S&P Dow Jones Indices LLC, S&P/Case-Shiller U.S. National Home Price Index [CSUSHPINSA]; S&P/Case-Shiller 20-City Composite Home Price Index [SPCS20RSA]; S&P/Case-Shiller 10-City Composite Home Price Index [SPCS10RSA].

Note: Percent change from a year ago, monthly, not seasonally adjusted.

not much above 3 percent.[2] As Shiller and Thompson note, that was a big miss given what we now know happened to prices. However, new buyers revised their expectations up sharply in 2021, nearly doubling across the four markets covered in the paper. That was still a big miss compared to realized price appreciation, but the relatively high volatility of one-year-ahead expectations is clearly evident. Long-run expectations responded more sluggishly, which is not surprising given the uncertainty around what the pandemic might have meant for housing markets. The paper's extensive discussion of written answers providing the context for reported expectations shows that about one-third of respondents in 2020 mentioned COVID-19 as a factor, but there was not a consensus on what it meant for house prices. By the time of the 2021 survey, there was widespread agreement that prices were rising in each of the four markets. As in previous cycles, survey respondents are well-informed about recent price movements

2. The 2020 survey was sent out in July of that year to buyers who closed on their homes in the first quarter of the year. Hence, they could have been searching and settled on the home in the latter part of 2019. This timing scenario also applies to the 2021 survey results discussed just below.

in their housing markets. The discussion of key events that buyers thought moved prices is illuminating, too—the combination of "low rates," "shortage of supply," and "high demand" suggests a supply-demand imbalance—at least in the short run.

Here, I think it would be useful for students of housing markets to try to bring in other data—on the supply side in particular. Hopefully, this could be done in next year's survey by adding supplementary questions. Supply was affected in two obvious ways. One was supply chain snafus which delayed builders from completing units and bringing them to the market. Government data indicate a meaningful increase in the number of months it took to complete a house once started. For example, 14 percent of all single-family units were completed within three months in 2019, versus only 9 percent in 2021. In general, it took longer to deliver a home to market in 2020 and 2021.[3] The other, and I think less well understood, supply side change involves the sharp drop in listings by existing owners. This latter phenomenon is important because buyers can purchase a new home or one from the existing stock. If nobody lists their home because, say, they were deterred by the need for social distancing or reluctant to relocate in uncertain times, then purchasers have to buy a new home. Whatever the causes, there were sharp declines in listings. One example comes from the Phoenix area. Data from an Arizona listings service indicate that the number of active listings fell from just over 18,000 in the first quarter of 2019 to just over 14,000 in the first quarter of 2020. And it got worse from there, as the analogous figure for the first quarter of 2021 was about 4,400.[4] Multiple listing services from different markets would have to be accessed to provide a clear national picture, but there seems little reason to believe that behavior in Phoenix was substantially more conservative than other metropolitan areas.

These negative shocks to the supply side of housing markets occurred simultaneously with an increase in the demand to be an owner versus a renter. US Census Bureau data show that the propensity to own jumped by 1.1 percentage points from the first quarter of 2019 to the same quarter in 2020 (from 64.2 percent to 65.3 percent) and then rose a bit further to 65.6 percent in 2021. There were nearly 130 million total households in the United States in 2020, so even a small percentage change implies a large

3. See US Census Bureau, "New Residential Construction," https://census.gov/construction/nrc/index.html, for more detail.

4. These figures come from the Arizona Regional Multiple Listing Service. They are aggregated from monthly data available at ARMLS, "Statistics," https://armls.com/statistics, and are for the Phoenix area, Maricopa County specifically.

absolute jump in purchasers.[5] Purchases of new home sales jumped sharply during the pandemic. The US Census Bureau reports that about 822,000 new homes were bought in the first year, 2020, versus only 683,000 in 2019, for an increase of 20 percent.[6] Existing home sales increased by a much smaller 6.3 percent between 2019 and 2020, from about 4.765 million units to 5.066 million units.[7] The sharp decline in listings noted above could have contributed to the relatively smaller increase in purchases out of the existing stock.

There seems much to learn from trying to put the expectations data in context, specifically with respect to perceived or actual changes in supply and demand. I agree with the statement in the text that COVID-19 heightened emotions and helped drive some people to find the "perfect house with space for meaningful new and different activities." The demand side matters for sure, and there is much interesting discussion in the paper about how this might relate to reported expectations. However, I encourage more exploration of the opportunity that the pandemic presents to study how the supply side may be affecting how people perceive housing markets. I suggest doing this through the survey's supplementary questions in 2022 and subsequent years. Perhaps the most obvious starting point would be to ask whether the respondents bought a new or existing home. I would also continue the question asked in the most recent survey about within-metropolitan area location of the home they purchased (i.e., central city versus suburb as in questions 33 and 34), as well as how much living area they desired (question 36). Finally, note that this is not an either-or issue, but one of "and," as in demand and supply.[8]

5. The homeownership rate data are taken from the *Quarterly Residential Vacancies and Homeownership* report (table 4), which updates quarterly; the report can be downloaded from the US Census Bureau, https://www.census.gov/housing/hvs/files/currenthvspress.pdf. The household data are from the Federal Reserve Bank of St. Louis FRED Economic Data, "Total Households," https://fred.stlouisfed.org/series/TTLHH.

6. The historical data can be downloaded from US Census Bureau, "New Residential Sales," https://www.census.gov/construction/nrs/historical_data/index.html. Purchases of newly constructed homes remained elevated in 2021 at 771,000 versus only 617,000 in calendar year 2018.

7. These data may be downloaded at National Association of Realtors, "Housing Statistics," https://cdn.nar.realtor/research-and-statistics/housing-statistics. The next year, 2021, saw a jump to 5.413 million purchases out of the existing stock, but that still is only a 7 percent increase over 2020.

8. In a related vein, I would even add supply side fundamentals to the demand side factors mentioned in question 17.2. That question asks whether the respondent believes more that psychological forces versus market fundamentals have had a stronger influence on recent house price trends. The market fundamentals—population growth, interest rates, and employment growth—all shift the demand schedule.

Another striking feature about this more recent era from figure 1 is how truly national the impact on prices has been. In the boom leading up to the global financial crisis, there was much more heterogeneity across markets. We all know about the so-called sand state market bubbles (Las Vegas, Phoenix, and Miami), but the Case-Shiller price series from other Sunbelt markets in North Carolina (Charlotte), Georgia (Atlanta), and Texas (Dallas) show no such boom during the run-up to the global financial crisis. It strikes me as worth considering why we do not see as much cross-market variation now. Changes in short- and long-term expectations already are compared and contrasted across four markets, and that obviously can be done in next year's survey, too. However, I would be interested in seeing a breakdown of written responses by market to learn whether different points were raised or whether buyers in, say, Milwaukee and San Francisco tended to reference similar or different factors. Finally, this is another area where coordination with other surveys could be valuable for our understanding of housing markets. There is much useful variation across the four markets surveyed here, but there is even greater variety across the country. Differences across markets, especially over time, provide a potentially useful context in which to study housing market behavior.

CONCLUSIONS This paper represents the continuation of a remarkably innovative research program into housing market expectations. The research it has spawned is now legion in scope and influence. Going forward, I hope we learn more about housing markets by exploiting the data in the written answers to supplementary questions that delve into issues beyond the survey respondents' specific views on future price appreciation. It would be useful to start publishing the underlying micro data on these answers, with an appropriate lag. Who knows what insights new textual analysis programs might provide? I also encourage more questions pertaining to the supply side of housing markets, particularly during and after the COVID-19 crisis.

REFERENCES FOR THE GYOURKO COMMENT

Case, Karl E., Robert J. Shiller, and Anne K. Thompson. 2012. "What Have They Been Thinking? Homebuyer Behavior in Hot and Cold Markets." *Brookings Papers on Economic Activity*, Fall, 265–98.

Kuchler, Theresa, Monika Piazzesi, and Johannes Stroebel. 2022. "Housing Market Expectations." Working Paper 29909. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29909.

GENERAL DISCUSSION Frederic Mishkin related his first question to a point made by Joseph Gyourko about Robert Shiller's "great disappointment"

of attempting to create settings where arbitrage can make housing markets more efficient. He explained that one of the important aspects of the authors' paper and earlier work on this survey is that housing markets are expected to be settings where rational expectations and market efficiency are unlikely to be found because of the lack of arbitrage. Mishkin wondered why this arbitrage continued to be elusive.

Mishkin's second question was posed to Gyourko and Adam Guren. He agreed with Guren's and Gyourko's assessments that the current rise in housing prices is driven by real factors rather than a traditional bubble; supply chain difficulties and increasing labor costs would cause housing prices to increase without a bubble forming. In this setting, a rise in home prices is efficient, since it comes from input costs and a corresponding increase in price increases the use of those inputs. Mishkin asked how much of the rise in housing prices was a public policy issue. He pointed out increasing supply restrictions from zoning laws over the last twenty years with an example from near his own home and asked how much of a role public policy played in increasing the difficulty of building new housing. He also asked why zoning laws have changed in the last twenty years.

Caroline Hoxby asked about heterogeneity of the period over which buyers forecast prices. For example, a house flipper may plan to own a house for only a short time, while others may plan to own and live in a house for decades, and still other homeowners may lie between those extremes. Depending on how long an owner plans to keep a house, their forecast of short- or medium-term price fluctuations may change. House flippers may not think in-depth about long-term price changes, while those who own their homes until retirement may not consider short-term price changes. Hoxby's concern is that the authors' data on expectations in price fluctuations look very smooth, which may be hiding a large amount of variation among homeowner expectations. Particularly those who plan to own their homes for a long time, she stated, may not think carefully about price changes even in the next ten years because they will not be selling their house in that period regardless of changes in the market. She wonders how the authors and discussants consider the period over which home buyers forecast prices in view of the home buyer's expected period of homeownership.

Gyourko addressed Mishkin's question about supply constraints in building new housing. He started by stating that he cannot answer this question with complete certainty, and much of that uncertainty stems from a lack of frequent, clear survey data. He added that this problem is worse in coastal markets, but regulation likely does not play a large role in constricting housing supply in cities like Nashville, Atlanta, and in most cities in Texas

(except Austin). Gyourko did not see Mishkin's example of homeowner communities preventing new housing development around their homes as representing a large segment of housing markets. He explained that while such actions are common, they do not affect the housing market as a whole because the number of new dwellings that are prevented is not large.

Guren discussed Hoxby's concerns about home buyer time horizons. He agreed that there may be heterogeneity among the period over which home buyers consider prices, and he stated that this heterogeneity is especially important in a bubble. He used this question to argue that the authors could maximize the impact of their research by publishing micro data from their survey, as the Federal Reserve Bank of New York survey does.[1] With this micro data, researchers could examine variation in home buyer expectations based on age, income, or other factors.

Shiller addressed the question of expectation time horizons first. He agreed that the opportunity for respondents to comment on housing markets provides rich context for the survey. Shiller proposed a focus group setting where skilled moderators speak with respondents about their expectations and can account for heterogeneity in the moment. This would better reflect the complexity of expectations than a fixed questionnaire like the authors' survey could.

Anne Thompson discussed Hoxby's question, pointing out that the authors' use of a trimmed mean reduced the effect of outliers on their conclusions. Because of this trimmed mean, heterogeneity in respondents' expectations for future home prices is limited to realistic forecasts. She stated that the individual responses, while they varied, were reasonable and rational, and no unrealistic outlying expectations distorted the data.

Janice Eberly brought up the large role of first-time home buyers and younger buyers in 2020 and 2021, pointing out that these groups were typically slow to enter the housing market before the onset of the pandemic. She asked whether the authors think that these groups represent pent-up demand and the effects of the pandemic, or whether other factors could explain the increasing role of first-time buyers and young buyers.

Thompson addressed this question, answering that while first-time and younger home buyers did see an increasing role before 2021, there was a large surge in the number of home buyers who fell into one of these

1. Federal Reserve Bank of New York, Center for Microeconomic Data, "Data Bank: Survey of Consumer Expectations," https://www.newyorkfed.org/microeconomics/databank.html.

groups in 2021. Individual responses show that first-time home buyers are older, on average, in 2021 than in earlier surveys. Thompson also discussed the increasing square footage of houses bought by first-time home buyers in 2021 compared to other years. These first-time home buyers, many of whom were millennials, had more savings because of high savings rates during the height of the pandemic among those who kept their jobs.