Janice Eberly and James H. Stock, Editors

Brookings Papers
ON ECONOMIC ACTIVITY

50TH Anniversary

SPRING 2020

STANSBURY and SUMMERS
on The Declining Worker Power Hypothesis: An Explanation for the Recent Evolution of the American Economy

ABRAHAM, HALTIWANGER, and RENDELL
on How Tight Is the US Labor Market?

PLAGBORG-MØLLER, REICHLIN, RICCO, and HASENZAGL
on When Is Growth at Risk?

ACEMOGLU, MANERA, and RESTREPO
on Does the US Tax Code Favor Automation?

DEL NEGRO, LENZA, PRIMICERI, and TAMBALOTTI
on What's Up with the Phillips Curve?

Brookings Papers ON ECONOMIC ACTIVITY

BPEA 50 SPRING 2020

BROOKINGS

BROOKINGS

www.brookings.edu

ISBN 978-0-8157-3876-3

Brookings Papers

ON ECONOMIC ACTIVITY

50TH Anniversary

SPRING 2020

JANICE EBERLY
JAMES H. STOCK
Editors

BROOKINGS INSTITUTION PRESS
Washington, D.C.

ISSN 0007-2303
ISBN: Print 978-0-8157-3876-3; ebook: 978-0-8157-3877-0

Brookings Papers
ON ECONOMIC ACTIVITY
50TH Anniversary

SPRING 2020

BPEA 50th ANNIVERSARY

The year 2020 marks the 50th anniversary of the *Brookings Papers on Economic Activity* (*BPEA*). Since its founding in 1970, *BPEA* has published high-quality independent research on timely economic challenges using state-of-the-art empirical methods and conceptual approaches. Currently led by coeditors Janice Eberly and James H. Stock, *BPEA* is widely recognized as a premier economics journal with a long-standing reputation for rigorous analysis and real-world application and relevance.

Ideas launched at *BPEA* often become policy soon afterward. In recent years, major findings have changed how we think about the student loan crisis, the high cost of health care, and long-term unemployment among American workers. *BPEA* research has also informed several of the Federal Reserve's most discussed tools and processes, including quantitative easing, forward guidance, and the current review of its monetary policy framework. *BPEA* addresses policy issues as they are emerging, such as the financial crisis in 2008, Brexit in 2016, and tax reform in 2017. It was a *BPEA* paper that identified the deepening mortality crisis due to "deaths of despair" in 2017.

For nearly half a century, *BPEA* has attracted top talent to serve as editors, authors, discussants, and advisers, including more than twenty Nobel laureates, Federal Reserve chairs, members of the President's Council of Economic Advisers, and chief economists from both public and private financial institutions. *BPEA* conferences have convened these distinguished economic experts and acted as a trusted platform for lively, intellectual debate and discussion. The need for the platform that the *Brookings Papers* creates for generating and debating policy-relevant research is as great now as it was 50 years ago, and we are committed to ensuring the continued success of this enterprise for the next 50 years.

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

We would like to thank the supporters of the *Brookings Papers on Economic Activity* conference and journal, including the Alfred P. Sloan Foundation; BlackRock Global Fixed Income; Brevan Howard Research Services Limited; General Motors Company; the National Science Foundation, under Grant No. 1756544; and State Farm Mutual Automobile Insurance Company.

CALL FOR PAPERS Most papers that appear in the *Brookings Papers on Economic Activity* are solicited by the editors, but the editors also consider submitted proposals. Editorial decisions are generally made about a year in advance of each conference. Therefore, proposals should be received by December 1 for the following fall conference and by June 1 for the following spring conference. Proposals can be submitted at http://connect.brookings.edu/submit-your-paper-to-bpea.

ACCESSING THE JOURNAL All past editions of the *Brookings Papers on Economic Activity*, including versions of the figures in color—along with appendix materials, data, and programs used to generate results—are made freely available for download at www.brookings.edu/bpea/search. To purchase subscriptions or single copies, visit www.brookings.edu/press, or contact the Brookings Institution Press at (866) 698-0010 or P.O. Box 465, Hanover, PA 17331-0465. Brookings periodicals are available online through both the Online Computer Library Center (contact the OCLC subscription department at (800) 848-5878) and Project Muse (http://muse.jhu.edu). Archived issues of the *Brookings Papers on Economic Activity* are also available through JSTOR (www.jstor.org).

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

Henry Aaron *Brookings Institution*
Stephanie R. Aaronson *Brookings Institution*
John Abowd *US Census Bureau*
Katharine G. Abraham *University of Maryland*
Daron Acemoglu *Massachusetts Institute of Technology*
George Akerlof *Georgetown University*
David Autor *Massachusetts Institute of Technology*
Martin Neil Baily *Brookings Institution*
Robert J. Barro *Harvard University*
Ben S. Bernanke *Brookings Institution*
Olivier Blanchard *Peterson Institute for International Economics*
William Brainard *Yale University*
Steven Braun *Council of Economic Advisers*
Ralph Bryant *Brookings Institution*
Gary Burtless *Brookings Institution*
Gerald Cohen *Haver Analytics*
Susan M. Collins *University of Michigan*
Charles Collyns *International Monetary Fund*
Richard N. Cooper *Harvard University*
Steven J. Davis *University of Chicago*
Jason DeBacker *University of South Carolina*
Marco Del Negro *Federal Reserve Bank of New York*
John Driscoll *Federal Reserve Board*
Karen Dynan *Harvard University*
Janice Eberly *Northwestern University*
Wendy Edelberg *Brookings Institution*
John G. Fernald *Federal Reserve Bank of San Francisco*
Benjamin M. Friedman *Harvard University*
Jason Furman *Harvard University*
Joseph Gagnon *Peterson Institute for International Economics*
William Gale *Brookings Institution*
Ted Gayer *Brookings Institution*
Mark Gertler *New York University*
Domenico Giannone *Amazon Inc.*
Tracy Gordon *Urban Institute*
Robert J. Gordon *Northwestern University*
Egor Gornostay *Peterson Institute for International Economics*
Daniel Goroff *Alfred P. Sloan Foundation*
Josh Gotbaum *Brookings Institution*
François Gourio *Federal Reserve Bank of Chicago*
Robert E. Hall *Stanford University*
John C. Haltiwanger *University of Maryland*
Thomas Hasenzagl *University of Minnesota*
Harry Holzer *Georgetown University*
Julie L. Hotchkiss *Federal Reserve Bank of Atlanta*
Caroline Hoxby *Stanford University*
Lawrence Katz *Harvard University*
Melissa Kearney *University of Maryland*
Aaron Klein *Brookings Institution*
Jeffrey Kling *Congressional Budget Office*
Michele Lenza *European Central Bank*
Daniel Lewis *Federal Reserve Bank of New York*
Nellie Liang *Brookings Institution*
Christos Makridis *Massachusetts Institute of Technology*

Andrea Manera *Massachusetts Institute of Technology*
N. Gregory Mankiw *Harvard University*
Mark Mazur *Urban-Brookings Tax Policy Center*
Jeffrey Miron *Harvard University*
Frederic Mishkin *Columbia University*
Robert A. Moffitt *Johns Hopkins University*
Ryan Nunn *Federal Reserve Bank of Minneapolis*
Christina Patterson *Northwestern University*
George Perry *Brookings Institution*
Jonathan Pingle *BlackRock*
Mikkel Plagborg-Møller *Princeton University*
James M. Poterba *Massachusetts Institute of Technology*
Giorgio E. Primiceri *Northwestern University*
Valerie Ramey *University of California, San Diego*
Richard Reeves *Brookings Institution*
Lucrezia Reichlin *London Business School*
Ricardo Reis *London School of Economics*
Lea E. Rendell *University of Maryland*
Pascual Restrepo *Boston University*
Giovanni Ricco *University of Warwick*
David H. Romer *University of California, Berkeley*
John Sabelhaus *Washington Center for Equitable Growth*
Ayşegül Şahin *University of Texas, Austin*
Claudia Sahm *Washington Center for Equitable Growth*
Brian Scholl *U.S. Securities and Exchange Commission*
Jay Shambaugh *Brookings Institution*
Louise Sheiner *Brookings Institution*
Christopher Sims *Princeton University*
Tara Sinclair *George Washington University*
David Skidmore *Brookings Institution*
Anna Stansbury *Harvard University*
Jeremy C. Stein *Harvard University*
Mark Steinmeyer *Smith Richardson Foundation*
James H. Stock *Harvard University*
Lawrence H. Summers *Harvard University*
Phillip Swagel *Congressional Budget Office*
François Velde *Federal Reserve Bank of Chicago*
Andrea Tambalotti *Federal Reserve Bank of New York*
Marcelo Veracierto *Federal Reserve Bank of Chicago*
David Wessel *Brookings Institution*
Justin Wolfers *University of Michigan*
Janet Yellen *Brookings Institution*
Eric Zwick *University of Chicago*

Francisca P. Alba *Brookings Institution*
Siobhan Drummond *Brookings Institution*
Stephanie K. Holzbauer *Brookings Institution*
Mary E. King *Brookings Institution*

ANNA STANSBURY
Harvard University

LAWRENCE H. SUMMERS
Harvard University

The Declining Worker Power Hypothesis: An Explanation for the Recent Evolution of the American Economy

ABSTRACT Rising profitability and market valuations of US businesses, sluggish wage growth and a declining labor share of income, and reduced unemployment and inflation have defined the macroeconomic environment of the last generation. This paper offers a unified explanation for these phenomena based on reduced worker power. Using individual, industry, and state-level data, we demonstrate that measures of reduced worker power are associated with lower wage levels, higher profit shares, and reductions in measures of the non-accelerating inflation rate of unemployment (NAIRU). We argue that the declining worker power hypothesis is more compelling as an explanation for observed changes than increases in firms' market power, both because it can simultaneously explain a falling labor share and a reduced NAIRU and because it is more directly supported by the data.

Since the early 1980s in the United States, the share of income going to labor has fallen, measures of corporate valuations like Tobin's q have risen, average profitability has risen even as interest rates have declined, and measured markups have risen. Over the same time period, average unemployment has fallen very substantially, even as inflation

Conflict of Interest Disclosure: Anna Stansbury is a doctoral candidate at Harvard University; Lawrence Summers is the Charles W. Eliot University Professor and president emeritus at Harvard University. Beyond these affiliations, the authors did not receive financial support from any firm or person for this paper or from any firm or person with a financial or political interest in this paper. They are currently not officers, directors, or board members of any organization with an interest in this paper. No outside party had the right to review this paper before circulation. The views expressed in this paper are those of the authors and do not necessarily reflect those of Harvard University.

has stayed low with no sign of accelerating—suggesting a decline in the non-accelerating inflation rate of unemployment (NAIRU).

We argue that the decline in worker power has been the major structural change responsible for these economic phenomena. A decline in worker power, leading to a redistribution of rents from labor to capital, would predict a fall in the labor income share; increases in Tobin's q, corporate profitability, and measured markups; and a fall in the NAIRU. In this paper, we estimate the magnitude of the decline in worker rent-sharing in the United States over recent decades, show that it is large enough to be able to explain the entire decline in the aggregate labor share and a substantial fraction of the decline in the NAIRU, and demonstrate that at the state and industry level, declines in worker power are consistent with changes in labor shares, unemployment, and measures of corporate profitability. Our focus on the decline in worker power as one of the major structural trends in the US economy is in line with a long history of progressive institutionalist work in economics, sociology, and political science, exemplified by Freeman and Medoff (1984), Levy and Temin (2007), Bivens, Mishel, and Schmitt (2018), Kristal (2010), Rosenfeld (2014), and Ahlquist (2017).

As an explanation for these recent macro trends, we believe that the evidence for the declining worker power hypothesis is at least as compelling as—and likely more compelling than—the other commonly posited explanations, specifically technological change, globalization, and rising monopoly or monopsony power.[1] While it is possible that globalization or technological change caused the decline in the labor share, it is difficult to reconcile each of these purely competitive explanations with the

1. For recent papers arguing that different aspects of globalization or technological change can explain the decline in the US labor income share, see, for example, Elsby, Hobijn, and Şahin (2013), Karabarbounis and Neiman (2014), Abdih and Danninger (2017), Autor and Salomons (2018), Acemoglu and Restrepo (2018), and Autor and others (2020). For papers arguing that rising monopoly power can explain the decline in the labor share or rising corporate valuations and markups, see Barkai (forthcoming), Gutiérrez and Philippon (2017, 2019), Eggertsson, Robbins. and Wold (2018), Farhi and Gourio (2018), and De Loecker, Eeckhout, and Unger (2020). For arguments that rising monopsony power could play a role in these trends, see Council of Economic Advisers (2016), Furman and Krueger (2016), Glover and Short (2018), Benmelech, Bergman, and Kim (2018), and Philippon (2020). For work on the role of the decline in worker power in the declining US labor share, see Elsby, Hobijn, and Şahin (2013) and Abdih and Danninger (2017), who both find some role for the decline in unionization but argue that it is not the dominant factor, and Kristal (2010) and Jaumotte and Osorio Buitron (2015), who argue that differential declines in worker power across countries can explain differential patterns of change in the labor share and income inequality.

rise in Tobin's q, average profitability, and measured markups over recent decades (which suggest an increase in economic rents accruing to capital owners). Alternatively, while it is possible that rising monopoly or monopsony power caused the decline in the labor share—and these would also be natural explanations for the rise in Tobin's q, average profitability, and measured markups—it is more difficult to reconcile rising monopoly or monopsony power with the decline in the NAIRU.

What do we mean by declining worker power? We consider the American economy to be characterized by three types of power, to varying degrees: monopoly power, monopsony power, and worker power. Firms' monopoly power—arising from explicit barriers to entry or from innate features of particular product markets, such as heterogeneous production technologies or short-run fixed costs—generates pure profits or rents. Firms' monopsony power in the labor market—arising from labor market concentration or labor market frictions—results in an upward-sloping labor supply curve to the firm, enabling the wage to be marked down to some degree below the marginal revenue product. Worker power—arising from unionization or the threat of union organizing, from firms being run partly in the interests of workers as stakeholders, or from efficiency wage effects—enables workers to increase their pay above the level that would prevail in the absence of such bargaining power.[2] This power gives workers the ability to receive a share of the rents generated by companies operating in imperfectly competitive product markets and can act as countervailing power to firm monopsony power.

In this framework, therefore, a decline in worker power results in a redistribution of product market rents from labor to capital owners.

What caused this decline in worker power? The decline in worker power in the US economy over recent decades was a result of three broad shifts. First, institutional changes: the policy environment has become less supportive of worker power by reducing the incidence of unionism and the credibility of the threat effect of unionism or other organized labor, and the real value of the minimum wage has fallen. Second, changes within firms: the increase in shareholder power and shareholder activism has led to pressure on companies to cut labor costs, resulting in wage reductions within firms and the fissuring of the workplace as companies increasingly outsource and subcontract labor.[3] And third, changes in economic conditions:

2. We use *worker power* as synonymous with bargaining power, rent-sharing power, and insider-outsider power.

3. For a detailed exposition of this trend, see Weil (2014).

increased competition for labor from technology or from low-wage countries has increased the elasticity of demand for US labor or, in the parlance of bargaining theory, has improved employers' outside option. In this paper, we emphasize the relative importance of the first two factors. While globalization and technological change surely did play some role in the decline in worker power, the cross-sector and cross-country evidence suggests that they are unlikely to have been the most important factors; within the United States, unionization has declined at similar rates across both tradable and non-tradable industries, and the decline in the US unionization rate has been much more pronounced than in many other countries (all exposed, to some extent, to similar international trends in technology and globalization).

We start our analysis in section I by examining the empirical evidence of a decline in worker power. Most notable is the decline of the private sector union membership rate, from over one-third at its peak in the 1950s to 6 percent today. In addition, the private sector union wage premium has declined somewhat since the early 1980s, suggesting that unionized workers are less able to share in the rents created by firms than they were in the past.

A different type of evidence of the importance of worker power comes from the fact that even without unions, workers may receive wage premia in other settings. Workers in larger firms and in certain industries (like manufacturing, mining, telecommunications, and utilities) receive substantially higher wages relative to observably equivalent workers in smaller firms or in other industries, and evidence suggests that these large firm and industry wage differentials to a large extent reflect rents. But workers' ability to receive rents in large firms or in high-rent industries appears to have declined. Using the Current Population Survey (CPS), we show that since the 1980s there has been a decline of about one-third in the large-firm wage premium and a decline of about one-third in the dispersion of industry wage premia.

A further source of evidence that worker power has been attenuated is the apparent decline in the relationship between workers' pay and the profitability, revenues, or product market power of their firm or industry. In a classically competitive labor market, workers' pay is determined by the marginal product of labor within their labor market, and there should be no correlation between a worker's pay and their firm's or industry's performance. In practice however, there is a positive relationship (suggesting a degree of rent-sharing). We show that the strength of this relationship has diminished over time: in manufacturing industries, the degree to which increases in revenue productivity translate into higher pay has declined since the 1960s, and we find suggestive evidence of a broad-based weakening in the relationship between industrial concentration and pay across sectors.[4]

So a large body of evidence points to a decline in worker power. But how big is this decline, in macroeconomic terms? In section II, we use our estimates of the union wage premium, large-firm wage premium, and industry wage premia to quantify the magnitude of the decline in total rents going to labor over 1982–2016. We demonstrate that labor rents are an important macroeconomic phenomenon and that they have declined substantially, from 12 percent of net value added in the nonfinancial corporate business sector in the early 1980s to 6 percent in the 2010s. (This is likely an underestimate, since we cannot quantify explicitly the decline in labor rents caused by the rise of activist shareholders.) This decline in labor rents is largely due to changes that have taken place within industries, rather than changes that have taken place across industries as employment has shifted from manufacturing to services.

The decline in labor rents could have been driven by either a destruction of rents available to be shared (as product market competition increased, perhaps as a result of globalization) or a redistribution of rents from labor to capital. Industry-level evidence tends to suggest that the decline in labor rents was largely a result of the latter: the majority of industries which saw substantial declines in rents to labor also saw substantial increases in profits to capital over 1987–2016.[5] And in manufacturing—the sector with the biggest decline in the labor share—the manufacturing industries with the greatest exposure to low-wage import competition were *not* the industries with the biggest declines in labor rents.

In section III, we demonstrate that the trends in factor shares, corporate profitability, Tobin's q, and measured markups that have sometimes been attributed to rising monopoly power can be equally or more convincingly

4. Note that this is a different issue than the one we addressed in Stansbury and Summers (2019). There we investigate the degree to which there is a relationship between changes in productivity and changes in compensation at the level of the whole economy. We find a nearly one-to-one relationship between changes in productivity and pay at the level of the whole economy over the postwar period, which has not attenuated since the 1970s and '80s. This finding could be consistent with either competitive or imperfectly competitive labor markets and is not inconsistent with our finding that the relationship between productivity and pay at the industry level has weakened (which indicates a decline in the degree of rent-sharing within different industries).

5. Our industry-level analysis spans 1987–2016, the longest period with data for consistent North American Industry Classification System (NAICS) industries.

explained by our hypothesis of declining worker power. We begin by replicating the recent decomposition exercise conducted by Farhi and Gourio (2018), who suggest that trends in factor shares, the profitability of capital, the investment-capital ratio, the risk-free rate, and other macroeconomic variables can be explained by an increase in average markups—alongside rising risk premia and increased unmeasured intangibles. In this framework, they estimate that average markups in the United States rose from 7 percent to 15 percent from the 1980s to the 2000s. While their analysis makes it clear that there are changes that cannot be explained by a perfectly competitive model, we note that there is essentially no way in their framework to distinguish between the rise in markups they posit (indicating a rise in monopoly or monopsony power) and a fall in worker power. Modifying their decomposition, we show that our hypothesis of declining worker power—holding markups constant—can explain the macro facts in the model equally well.

Next, we take our measure of the magnitude of lost labor rents (calculated in section II from union wage premia, large-firm wage premia, and industry wage premia) to the aggregate data on the nonfinancial corporate sector. We show that our estimate of the decline in labor rents—at roughly 6 percent of nonfinancial corporate sector value added since the 1980s—is big enough to (over)explain the entire decline in the net labor share. At the state level, our measure of the decline in the labor rent share is predictive of changes in the labor share over 1984–2016.

We then compare trends in labor rents, labor shares, profitability, and measures of Tobin's q for fifty-one industries.[6] We show that industries with larger declines in labor rents over 1987–2016 had much larger declines in their labor shares and increases in their average profitability. In horserace regressions, industry-level labor rents have substantially more power to explain changes in labor shares, profitability, and Tobin's q than measures of product market concentration (which have been used as indicators of a rise in monopoly power).

In section IV, we argue that the decline in worker power would be consistent with another highly salient aspect of the macro experience of recent decades: the substantial decline in both average unemployment and average inflation. The unemployment rate was below 5 percent, the level previously thought to have been the NAIRU, for nearly half of the twenty-three years

6. Where industries are defined at roughly the three-digit level of the NAICS.

from 1997 to 2019 and was at or below 4 percent from May 2018 until February 2020, at levels not reached since the 1960s. At the same time, inflation has been low and has shown little sign of accelerating. These facts suggest that there has been a quite substantial decline in the NAIRU or a flattening of the Phillips curve or both.

Almost all models of declining worker power predict a fall in the NAIRU, as the decline in the cost of labor increases firms' hiring or as wait unemployment falls. In keeping with these predictions, we show that states and industries with bigger falls in worker power over the last four decades saw bigger falls in their unemployment rate. Extrapolation from our analysis of state-level unemployment rates suggests that the aggregate change in worker power could be big enough to explain a large fraction of the decline in the NAIRU. (We further verify this conclusion with informal calculations in online appendix E, drawing on various models of the relationship between worker power and the NAIRU.)[7] We note, on the other hand, that an increase in monopoly power offers no explanation for the decline in the NAIRU. If anything, it has usually been thought to act in the other direction: in the presence of downward nominal wage rigidity, rising monopoly power would tend to predict rising prices (as firms transition to a new equilibrium of higher markups and higher prices) alongside a rise in unemployment (as the rise in monopoly power leads to a restriction in output). Increasing monopsony power would tend to be associated with less, rather than more, hiring and so does not provide a natural explanation for a declining NAIRU. And globalization and technological change, while possibly disinflationary, would tend to increase average unemployment by increasing disruption and structural change in the economy, making their implications for the NAIRU ambiguous.

In section V, we address possible objections to the declining worker power hypothesis. First, we show that the apparent weakness of investment relative to fundamentals—which has been a major motivator of the monopoly power argument—can be reconciled with our hypothesis. Second, we show that recent research emphasizing the importance of between-firm reallocation in explaining changes in factor shares is consistent with the declining worker power hypothesis. Third, we note that our measure of labor rents does not incorporate any increase in labor rents which may have accrued to the highest earners—such as executives or top

7. The online appendixes may be found at the *Brookings Papers* web page, www.brookings.edu/bpea, under "Past BPEA Editions."

earners in finance—and should be thought of as a measure of the decline in the rents accruing to the *majority* of workers.[8] Fourth, we argue that the rise in occupational licensing has likely not played a major role in the trend in aggregate labor rents over recent decades. Finally, we note that the decline in the labor share has been much more pronounced in the United States than other industrialized economies similarly exposed to globalization and technological change, and we note that the decline in the labor share has been most pronounced in US manufacturing, which (given increasing globalization) is not an industry where a large rise in monopoly power seems likely to have occurred. We also note that there is little evidence of any large increase in import-adjusted sales concentration in manufacturing or in local-level sales concentration in services and that local labor market concentration has declined over time. Together, these suggest to us that globalization, technological change, or rising monopoly or monopsony power alone lack the ability to explain recent economic developments in a unified way.

While the focus of this paper is on the distribution of rents between labor and capital, we note that the decline of labor rents has also likely increased inequality in labor incomes: the declines in unionization and the real value of the minimum wage and the fissuring of the workplace affected middle- and low-income workers more than high-income workers, and some of the lost labor rents for the majority of workers may have been redistributed to high-earning executives (as well as capital owners). Consistent with these hypotheses, we show that the decline in labor rents was larger for non-college-educated workers than for college-educated workers, and we estimate, in a back-of-the-envelope exercise, that the decline in labor rents could account for a large fraction of the increase in the income share of the top 1 percent over recent decades.

Overall, we conclude that the decline in worker power is one of the most important structural changes to have taken place in the US economy in recent decades. Our emphasis on the decline of worker power is justified both by the strength of the direct evidence and by its ability to provide a unified explanation for a variety of macroeconomic phenomena: changes in labor and capital incomes, profitability, and the NAIRU.

This raises important challenges for policy. If a major feature of the US economy were a rise in monopoly or monopsony power, reducing

8. This is because our measure of labor rents is estimated in the CPS, which is top-coded for high earners and has higher nonresponse rates for these groups.

restrictiveness and increasing competition in markets could improve *both* efficiency and equity. But if, as we argue, the major explanation of the decline in the labor share and rise in corporate profitability is a decline in worker power, then measures to restrict monopoly or monopsony power alone—or indeed, to restrict globalization or technological change—may do little to reverse this trend. More profoundly, if markets are innately characterized by some degree of imperfect competition and rents, then completely eliminating all sources of market power may not be feasible. Instead, if increases in the labor share are to be achieved, institutional changes that enhance workers' countervailing power—such as strengthening labor unions or promoting corporate governance arrangements that increase worker power—may be necessary (but would need to be carefully considered in light of the possible risks of increasing unemployment).

I. Evidence of Declining Rent-Sharing in US Labor Markets

Why do firms share rents with workers? There are three groups of reasons. First, workers may be able to lay claim to rents directly, as a result of either explicit bargaining power through unions or implicit bargaining power through the threat of union organizing (Freeman and Medoff 1984) or via some other ability to wield power within the firm. Second, some firms may be run partly in the interests of workers as stakeholders rather than solely in the interests of shareholders. Third, it may be in firms' interests to share rents with workers for efficiency wage reasons—where workers are paid an above-market wage to incentivize effort (Yellen 1984)—or to maintain morale, perhaps as a result of fairness norms, as discussed in Akerlof and Yellen (1986).[9] Efficiency wages may also play a role in reducing the cost to firms of paying above-market wages: if worker productivity increases when wages rise, then some of the extra cost of sharing rents with workers is offset by productivity benefits (Bulow and Summers 1986; Summers 1988).

Evidence from a wide range of sources has demonstrated the existence of rent-sharing in the US labor market. Unionized workers, workers at certain firms (particularly at large firms), and workers in specific industries

9. The rents received by workers may be true rents or pure profits generated by a firm's monopolistic power in the product market—or they may be quasi rents generated by sunk investments (Grout 1984; Caballero and Hammour 2005) or by the cost of recruiting new workers either in a frictional labor market or in a setting where job-specific training is required (Mortensen and Pissarides 1999; Manning 2003).

receive substantial wage premia relative to observably equivalent workers. Similar wage premia also exist for workers who switch jobs, suggesting they do not reflect unobserved worker characteristics. These wage premia tend to be positively correlated with indicators of rents at the firm and industry levels, including profits and concentration, and inversely correlated with quit rates (both of which are suggestive of rent-sharing). In addition, there is evidence of sizable pass-through of industry- or firm-level shocks to productivity and profits into workers' compensation. And there is a large body of work documenting persistent wage losses for displaced workers, which partly reflect lost rents.[10]

Over recent decades however, a number of forces have likely reduced labor rents in the United States, particularly for lower-wage workers. Most obvious have been the decline in unionization and union bargaining power and the erosion of the real value of the minimum wage. In addition, the increase in shareholder activism and the rise of the shareholder value maximization doctrine increased the power of shareholders relative to managers and workers, likely increasing pressure on firms to cut labor costs and, in particular, to redistribute rents from workers to shareholders.[11] The increased fissuring of the workplace, with outsourcing of noncore business functions, may be an outgrowth of this phenomenon (Weil 2014). In this section, we present a range of empirical evidence of this decline in rent-sharing.

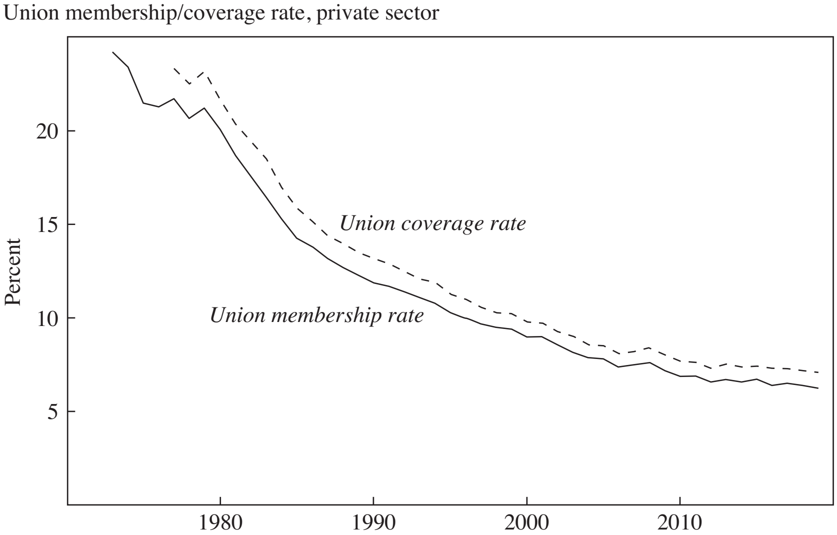
I.A. Declining Unionization Rates

Unions are the most clear-cut example of workers having rent-sharing power. Unionized workers receive significantly higher wages than observationally equivalent nonunion workers, with most estimates of the private

10. We briefly review evidence on union, industry, and firm size wage premia later in this section. For evidence on firm-specific wage premia, see Groshen (1991), Davis and Haltiwanger (1991), and the extensive literature estimating "AKM" models, starting with Abowd, Kramarz, and Margolis (1999). Estimates from the AKM literature suggest that firm effects and the covariance of worker and firm effects can explain 17–20 percent of the variance of wages (Abowd, Lengermann, and McKinney 2003; Abowd, McKinney, and Zhao 2018; Song and others 2019) and that around one-third of this reflects rents (Sorkin 2018). For evidence on wage losses for displaced workers, see Jacobson, LaLonde, and Sullivan (1993), Davis and von Wachter (2011), and Lachowska, Mas, and Woodbury (2018), among others.

11. See, for example, Shleifer and Summers (1988), who argue that a primary effect of hostile takeovers is the redistribution of value to shareholders from other stakeholders. Some evidence consistent with this mechanism can be found in Davis and others (2019), who find that wage premia in target firms were largely erased after private equity buyouts.

Figure 1. Union Membership and Coverage Rates, Private Sector



Source: Union Membership and Coverage Database from the CPS, constructed by Barry Hirsch and David Macpherson, www.unionstats.com.

sector union wage premium between 15 percent and 25 percent (Rosenfeld 2014).[12] But the ability of workers to share in rents through unions has declined substantially in recent decades. Private sector union membership gradually declined from a peak of around one-third in the 1950s to 24 percent in 1973 and then declined more rapidly, reaching 6 percent in 2019 (figure 1).[13] In addition, estimates of the union wage premium suggest that it has declined since the early 1980s.[14]

12. Empirical evidence is consistent with this wage premium representing a redistribution of rents from capital to labor. For example, Abowd (1989) finds substantial evidence to support a dollar-for-dollar trade-off between workers and shareholders in union contract settlement data. Lee and Mas (2012) show that new unionization reduces firms' equity value. If this represents a redistribution of rents from capital to labor, the magnitude of the average effect they find would be consistent with a 10 percent union wage premium.

13. The measured decline in the unionization rate may be an underestimate: as the unionization rate approaches zero, misclassification bias tends to produce inflated estimates (Card 1996; Western and Rosenfeld 2011).

14. We estimate the union log wage premium for private sector workers in the Current Population Survey, Outgoing Rotation Group (CPS-ORG), regressing the log hourly wage on a dummy variable for union membership or coverage with controls for education, demographics, geography, occupation, and industry (see online appendix A.1 for more details). Our estimate falls from 21 log points in 1982 to 15 by 2019. These are both within the historical range over the twentieth to twenty-first century as estimated by Farber and others (2018).

Note that the impact of unions on workers' ability to receive rents likely extended beyond the workers who were unionized receiving wage premia. In industries where pattern bargaining was common, nonunionized firms would match the wage increases in union contracts (with the most famous example being the 1950 Treaty of Detroit). Even without pattern bargaining, the threat effect of unionization of workers in nonunion firms likely incentivized firms to offer better wages and benefits than they otherwise would have (Leicht 1989; Farber 2005; Denice and Rosenfeld 2018).[15] And union bargaining power may have more generally supported norms of equity in pay structures (Western and Rosenfeld 2011).

The decline in unionization rates and union bargaining power was driven by a combination of institutional factors, which weakened labor law and its enforcement, and economic factors, which increased the elasticity of demand for labor and so weakened workers' ability to bargain for higher wages. Institutional factors included the breakdown of pattern bargaining in the 1980s, the expansion of the number of right-to-work states, and decreasing political support for and enforcement of labor laws.[16] Economic factors that reduced worker bargaining power included increased import competition for manufactured goods and deregulation of transportation and telecommunications, both of which reduced firms' ability to compete while paying high wages (Peoples 1998; Levy and Temin 2007; Rosenfeld 2014). Note, however, that these economic factors are unlikely to have been the main drivers in the decline in US unionization; the proportional decline in the unionization rate from the mid-1980s to the mid-2000s was almost identical across a range of sectors which had very different exposures to globalization, technological change, and

15. Unions may also raise wages for nonunion workers in frictional labor markets as employers raise wages to retain the ability to hire easily (Manning 2003). On the other hand, unions may have negative spillovers on the wages of nonunion workers if the union raises wages but restricts employment in the union sector (Oswald 1982). Overall, though, evidence suggests a positive correlation between unionization rates and nonunion wages, suggesting that union spillovers are on net positive (Farber 2005; Leicht 1989; Neumark and Wachter 1995; Denice and Rosenfeld 2018; Fortin, Lemieux, and Lloyd 2018).

16. See, for example, Levy and Temin (2007) and Rosenfeld (2014). Workers' ability to organize was reduced both by a direct weakening of labor law and labor law enforcement and by an increased corporate use of union avoidance tactics (Bronfenbrenner 2009; McNicholas and others 2019). The fissuring of the employment relationship has also decreased workers' ability to organize: workers employed as independent contractors or in franchises often have their terms of employment to some extent dictated by the end employer or franchisor (respectively) but lack the legal ability to collectively bargain with that end employer (Paul 2016; Steinbaum 2019).

deregulation over the period in question (manufacturing, mining, transportation and utilities, retail trade, construction, and wholesale trade), and the rate of unionization has declined much more quickly in the United States than in most other industrialized economies, despite similar trends in globalization and technology (Schmitt and Mitukiewicz 2012; Denice and Rosenfeld 2018).[17]

I.B. Declining Large-Firm Wage Premium

A large body of literature shows that large firms pay workers higher wages than their otherwise equivalent counterparts at smaller firms.[18] While this firm size effect could be driven by a number of different causes—workers with higher unobserved productivity, compensating differentials, a greater propensity to pay efficiency wages, a decision to pay higher wages to fill vacancies faster—several studies have found that even when attempting to account for these possibilities a large unexplained firm size premium often remains (Brown and Medoff 1989). This implies that some substantial portion of the large-firm wage premium reflects rents to labor.[19] Over recent decades, however, the large-firm wage premium has fallen (Hollister 2004; Even and Macpherson 2012; Cobb and Lin 2017; Song and others 2019). Estimating the large-firm wage effect for observably equivalent private sector workers over 1990–2019 from the CPS Annual Social and Economic Supplement (CPS-ASEC), we find a substantial decline in wage premia for workers at firms with 500 or more employees, relative to workers at small firms (figure 2), likely indicating a decline in rent-sharing.[20] (To interpret it as something *other than* a decline in rent-sharing, there must have been either a substantial reduction in compensating differentials as small firms became relatively worse to work at or

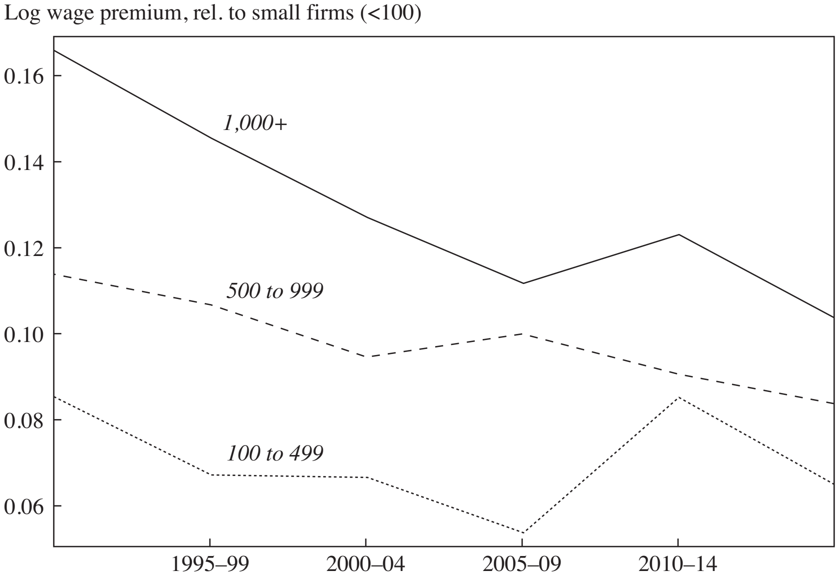
17. See online appendix C.1 for unionization rates by industry. Note also that while Acemoglu, Aghion, and Violante (2001) argue that the decline of unionization was endogenous, driven by skill-biased technological change, Farber and others (2018) find that the pattern of decline of US union membership is unlikely to be consistent with this.

18. See Brown and Medoff (1989), Bulow and Summers (1986), and Davis and Haltiwanger (1996).

19. This is consistent with large firms being more likely to have product market power—and so, rents.

20. We run log wage regressions on dummies for firm size and various demographic, occupation, and location controls. We obtain estimated for the firm size wage effects for workers at firms of 1000+, 500–999, and 100–499 workers, relative to firms with <100 workers. We regress on five-year pooled samples as the sample size is too small for precise annual estimates. See online appendix A.2 for more details.

Figure 2. Large Firm Wage Effect, Private Sector



Sources: Current Population Survey, Annual Social and Economic Supplement; authors' calculations.
Note: The large firm wage premium is estimated for firms with 100–499, 500–999, and 1,000+ employees for five-year periods over 1990–2019, controlling for education, demographics, geography, occupation, industry, and union status. More details on estimation procedures are in the text and in the online appendix A.2.

a reduction in the sorting of highly productive workers into large firms.) Note that if large firms' monopoly power had systematically increased over recent decades without any change in worker rent-sharing power, the large-firm wage premium would have been expected to increase rather than decrease.

I.C. Declining Variance of Industry Wage Differentials

A large body of work on the interindustry wage structure, over several decades, has found substantial and persistent dispersion of wages across industries for observably similar workers. Evidence suggests that industry wage differentials to a large extent reflect rent-sharing with workers: the wage differentials persist even when accounting for worker productivity differences and compensating differentials, and they are correlated with industry-level profitability, concentration, and capital-labor ratios (Dickens and Katz 1987; Krueger and Summers 1988; Katz and Summers 1989; Gibbons and Katz 1992; Abowd and others 2012).[21]

Using the Current Population Survey, Outgoing Rotation Group (CPS-ORG), we estimate industry wage differentials for private sector workers in each year over 1984–2019. We regress log wages on a set of industry dummies at different levels of aggregation (eighteen sectors, seventy-seven industries, or 250 detailed industries) alongside controls for education, demographics, geography, occupation, and union membership or coverage.[22] This gives us a set of estimated wage fixed effects for each industry. If rent-sharing with labor has declined in recent decades, we would expect the variance of industry wage premia to have declined. As figure 3 shows, this is the case at all levels of industry aggregation.[23]

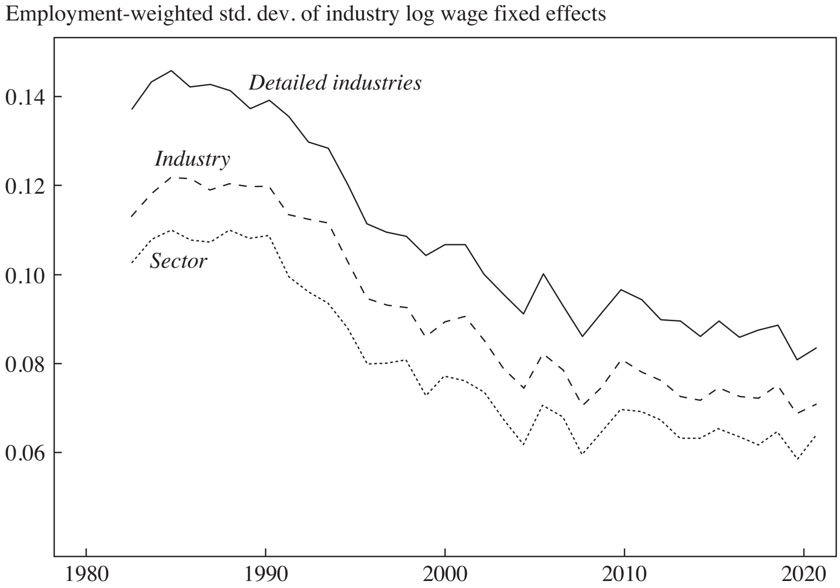
As with the decline in firm size wage effects, it is possible that the decline in the variance of industry wage effects was not a result of falling rent-sharing but was instead a result of changing compensating differentials or sorting by unobserved worker productivity. We have, however, no a priori reason to believe that there has been a substantial change in compensating differentials in the necessary direction (as it would imply that high-wage industries used to have much worse amenities but have improved over time). We can test the sorting explanation by estimating

21. Further evidence that these premia indicated the presence of rents included the fact that wage premia for workers in different occupations in the same industry were highly correlated and that industries with higher wage premia tended to have lower quit rates and higher ratios of applicants to job openings, shown in the previously mentioned studies, as well as Slichter (1950), Ulman (1965), and Holzer, Katz, and Krueger (1991). More recently, Abowd and others (2012) found that industry wage differentials were strongly correlated with firm effects in an AKM decomposition, strengthening the case that they are to some extent a function of rents.

22. These sectors correspond to NAICS sectors, the industries correspond to Bureau of Economic Analysis (BEA) industry codes (roughly NAICS three-digit), and the detailed industries correspond to Standard Industrial Classification (SIC) industries. More details on estimation are in the online appendix A.3. Note that the CPS-ORG data are top-coded, so we will not observe changes in firm size or industry wage premia for very high earners.

23. Kim and Sakamoto (2008) also find evidence of a decline in interindustry wage dispersion using the CPS-ORG, albeit with a different methodology. Note that our result does not conflict with the result of Haltiwanger and Spletzer (2020), who find that the dispersion of average log earnings across industries has risen over 1997–2013; this pattern also exists in our raw CPS data but is reversed once occupation and individual characteristics are controlled for. In addition, much of the decline in industry wage differentials we identify in the CPS occurs before 1997.

Figure 3. Standard Deviation of Industry Wage Effects



Sources: Current Population Survey, Outgoing Rotation Group; authors' calculations.
Note: Industry fixed effects are calculated as the fixed effect on industry dummies in annual log wage regressions over 1984–2019, with demographic, location, and occupation controls. "Sector" refers to eighteen aggregated NAICS sectors, "Industry" refers to seventy-seven industries (roughly NAICS three-digit level), and "Detailed industries" refers to 250 SIC industries. More details on estimation procedures are in the text and in the online appendix A.3.

industry fixed effects using the longitudinal component of the CPS, which enables us to control for worker-level unobserved productivity. The proportional decline in the variance of industry fixed effects estimated longitudinally is as large as for the cross-sectional estimates, suggesting that the decline we observe is *not* driven primarily by a change in the degree of sorting of highly productive workers into high-wage industries.[24]

24. More details on the longitudinal estimates are available in the online appendix A.4. Note also that even to the extent that industry fixed effects *do* represent rents, a decline in the dispersion of industry fixed effects could be a result of a decline in the dispersion of industry-level rents, holding constant the degree of rent-sharing. This does not appear to be the case: the cross-industry dispersion of various measures of profitability has not fallen over the period. Another possibility is that the fall in the employment-weighted standard deviation of industry fixed effects simply represents a reallocation of workers from high-rent to low-rent industries. This also does not appear to be driving the result: the non-employment-weighted standard deviation of industry fixed effects has fallen by roughly the same amount.

I.D. Decreased Pass-Through of Productivity and Profit Shocks

A different source of evidence that worker power has been attenuated is the apparent decline in the relationship between workers' pay and the profitability, revenues, or product market power of their firm or industry. A perfectly competitive labor market would imply no relationship between firm- or industry-level performance and workers' pay, but in practice there is substantial evidence that firms and industries with higher productivity or profitability do pay more to observably equivalent workers, as reviewed in Card and others (2018).[25]

There is some evidence to suggest, however, that this relationship has weakened over time. Using the National Bureau of Economic Research and the US Census Bureau's Center for Economic Studies (NBER CES) manufacturing data, which covers 473 NAICS six-digit manufacturing industries over 1958–2011, we regress the annual change in log value added per worker on the annual change in log compensation per worker.[26] We find evidence of rent-sharing over the period: in years with 10 log points higher value added per worker, average pay in a given industry was 2.5 log points higher. But the strength of that relationship fell by about half from the 1960s and '70s to 2000–2011 (figure 4). In similar work, Bell, Bukowski, and Machin (2019) find a declining relationship between profits per worker and compensation per worker in US manufacturing industries, also using the NBER CES data. Benmelech, Bergman, and Kim (2018) report a decline in the relationship between output per hour and compensation per hour at the plant level in US manufacturing over 1978–2007. Together, this evidence is strongly suggestive of a decline in rent-sharing in US manufacturing: workers in firms and industries with higher revenue productivity and higher profits appear to share in this less than they used to.

We also examine evidence on the relationship between product market concentration and wages. At the very aggregated sector level, we find a positive relationship between average product market concentration and the sector wage premium, but the strength of that relationship declined

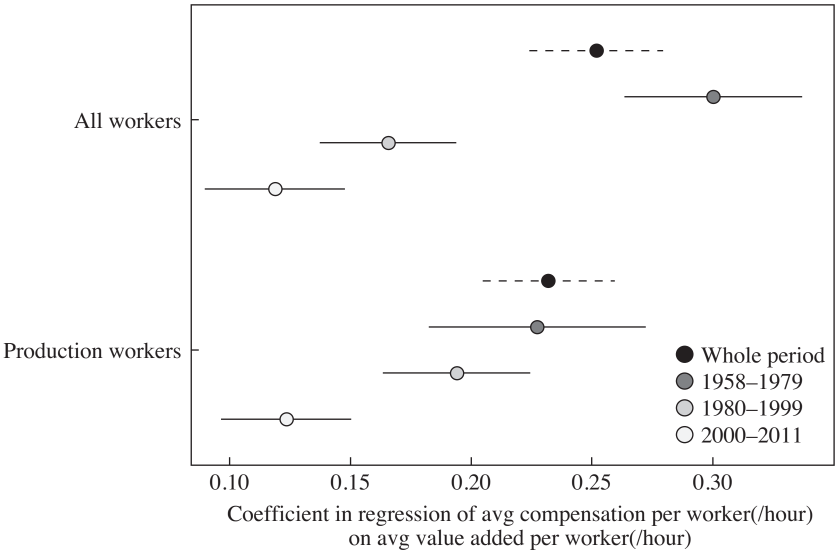
---

For more details, see the online appendixes A.6 and A.7. A further indication that our measure of industry labor rents *is* picking up rents is that we find that industries with higher wage premia have substantially and significantly lower quit rates, as found also by Holzer, Katz, and Krueger (1991).

25. See also the online appendix D.2 for a review of some of this evidence.

26. Following Stansbury and Summers (2019) we use a three-year moving average of each variable in the regression. Our results are robust to the choice of moving average length. Note that NAICS six-digit manufacturing industries are very narrowly defined: for example, NAICS 337110 "Wood kitchen cabinet and countertop manufacturing."

Figure 4. NAICS Six-Digit Industry-Level Regression of Average Compensation per Worker on Average Value Added per Worker, Manufacturing



Sources: NBER CES Manufacturing database; authors' calculations.

Note: Figure shows coefficients from manufacturing industry regressions: 1958–2011. "All workers" regresses the change in log compensation per employee on the change in log value added per employee, three-year moving averages, following the specification in Stansbury and Summers (2019). "Production workers" regresses the change in log average hourly production worker pay on the change in log value added per production worker hour, three-year moving averages. Regressions have NAICS six-digit industry and year fixed effects. Standard errors are clustered at the NAICS six-digit industry level. Dots represent point estimate and lines represent the 95 percent confidence interval. Each line and dot is from a separate regression.

over 1982–2012. In regressions of product market concentration on wage premia at the industry level, the general trend also suggests a weakening of the relationship between average concentration and the wage premium, but the change over time is not statistically significant.[27]

I.E. Increased Use of Domestic Outsourcing and Subcontracting

A final indicator that rent-sharing has declined is the increase in the use of outsourcing, subcontracting of business functions, and franchising; the

27. See online appendix C.2 for details of these analyses. Note that if product market concentration became a noisier measure of monopoly power over time, we might expect to see a weakening relationship between concentration and wage premia *even if* the underlying relationship between monopoly power and wage premia remained constant.

growth in independent contracting and the gig economy; and the decline in internal labor markets, often referred to jointly as the "fissuring" of the workplace (Weil 2014; Bernhardt and others 2016; Bidwell and others 2013). If workers' ability to share in firm-level rents depends on them being employed within the firm, then one would expect that this fissuring would lead to wage decreases, particularly for workers working (indirectly) for high-rent firms.[28] There is increasing evidence that outsourced workers receive wage penalties and that this is related to a loss of rents.[29] While the scale of fissuring is difficult to measure with existing data (Bernhardt and others 2016), evidence suggests that it is widespread. Weil (2019) estimates that—as a rough lower bound—19 percent of private sector workers were in industries where fissured arrangements predominate. Looking at specific occupations, the share of workers in security, cleaning, and logistics occupations who work in business services industries rose from less than 10 percent in 1970 to 35 percent, 25 percent, and 20 percent respectively in 2015 (Dorn, Schmieder, and Spletzer 2018).

II. Estimating the Magnitude of the Decline in Labor Rents

The evidence in section I paints a picture of declining rent-sharing with labor—but was it big enough to explain the macro trends we have seen? We use a back-of-the-envelope approach to estimate the total quantity of

28. Factors driving the fissuring of the workplace may have been an increase in shareholder pressure to cut labor costs, increased ability to coordinate and monitor the performance of contracted workers, increased focus on firm "core competencies," declining union presence, and an erosion of antitrust standards prohibiting nonprice vertical restraints (Weil 2014; Bernhardt and others 2016; Bidwell and others 2013; Steinbaum 2019). Factors which make rent-sharing more likely if workers are employed within the boundaries of the firm include the degree to which rent-sharing is determined by unionization or the threat of unionization and the degree to which rent-sharing depends on a sense of pay equity or internal labor markets within the firm.

29. Dube and Kaplan (2010) find that outsourced janitors and guards lose wage premia, consistent with a loss of firm-specific rents; Dorn, Schmieder, and Spletzer (2018) find evidence of a loss of wage premia for outsourced workers in food, cleaning, security, and logistics occupations; Mishel (2018) links the decline in the manufacturing wage premium to the increase in the use of staffing agencies; and Wilmers (2018) finds that workers at supplier firms which become dependent on a dominant buyer lose wages, consistent with a loss of rents. Evidence from Handwerker (2018) and Song and others (2019) is also consistent with the fissuring of the workplace leading to a loss of rents: Handwerker (2018) finds that wages are lower in firms with more concentrated occupational employment, and this concentration has increased over time; Song and others (2019) find an increase in the sorting of highly paid workers into high-paying firms (and vice versa).

labor rents in the US nonfinancial corporate sector for each year from 1982 to 2016, as follows:

$$\text{Total labor rents} = \text{union rents} + \text{industry rents} + \text{firm size rents}$$

where "union rents" refers to rents arising from union wage premia for unionized workers, "industry rents" refers to rents arising from industry wage premia, and "firm size rents" refers to rents arising from large-firm wage premia. We calculate union rents, industry rents, and firm size rents from our estimates of union, industry, and firm size wage premia as outlined below.[30] Note that our estimate is of the total quantity of labor rents for the *majority* of workers, excluding the very highest earners, since top-coding and nonresponse in the CPS mean we cannot estimate union, industry, or firm size rents for these earners.

II.A. Union Rents

For each year t, we estimate the share of total compensation in the nonfinancial corporate sector that was union rents, using estimates of the union log wage premium uwp_t, the union coverage rate in each year ucr_t, and compensation in the nonfinancial corporate sector, as follows:[31]

$$\text{Union rents}_t = \text{compensation}_t\left(1 - \frac{1}{1 + ucr_t\,(e^{uwp_t} - 1)}\right)$$

II.B. Industry Rents

For each industry j and year t, we estimate the share of total compensation in that industry that was industry rents. We start with our estimated industry fixed effects from log wage regressions, at the level of nineteen NAICS sectors for 1987–2016 and nine SIC sectors for 1982–1986. To calculate the industry wage premia from the estimated fixed effects, we first rescale the estimated industry fixed effects relative to the large industry

30. Full details of the calculation are in online appendix B.1. We focus on the nonfinancial corporate sector for our baseline estimates and present estimates of labor rents for the full corporate sector in online appendix B.2.

31. We estimate the union log wage premium from the CPS-ORG for 1984–2019 and use estimates from Blanchflower and Bryson (2004) for years 1982 and 1983. We estimate the union coverage rate for workers in private industries excluding finance, insurance, and real estate for 1984–2019 from the CPS-ORG and extend these back to 1982 using data on the private sector union coverage rate from unionstats.com.

with the lowest fixed effect, which is retail trade. (This calculation assumes that there are zero labor rents on average for workers in retail trade.) We then treat *half* of the deviation of each industry's fixed effect from the retail trade fixed effect as an industry wage premium ("rents"). We only consider half of the industry wage differentials to represent rents because, even though we have controlled for as many person-level characteristics as we can, there may still be worker sorting into industries on unobserved productivity differences and because part of the estimated interindustry wage differentials may reflect compensating differentials. While we choose simply to cut industry wage effects in half for transparency, we have reason to believe this is reasonable: first, our estimates of industry wage premia from the longitudinal component of the CPS, controlling for person-level fixed effects, are very highly correlated with our cross-sectional estimates and are exactly half as big on average; and second, we benchmark our estimates against estimates of industry wage premia and the degree of rent-sharing from two papers (Abowd and others 2012; Sorkin 2018) which use the AKM estimation method developed by Abowd, Kramarz, and Margolis (1999).[32] This approach gives us an estimated industry wage premium $iwp_{j,t}$ for each industry j, allowing us to calculate industry rents as:

$$\text{Industry rents}_t = \sum_{j}^{\text{industries}} \text{compensation}_{j,t}\left(1 - \frac{1}{e^{iwp_{j,t}}}\right)$$

where "compensation" refers to our estimate of total nonfinancial corporate sector compensation for each industry.[33]

II.C. Firm Size Rents

For each firm size class s and year t we estimate the share of total nonfinancial corporate compensation that was firm size rents, using our

32. For our benchmarking procedure, we take estimates for the average firm fixed effect across different US sectors over 1990–2001 from Abowd and others (2012) and apply Sorkin's (2018) estimate that one-third of firm fixed effects on average represent rents. This gives us a rough estimate of the average log wage premium due to rents in each sector, over 1990–2001. More details on our longitudinal fixed effect estimates are in online appendix A.4 and on our benchmarking procedure in online appendix A.5.

33. This is calculated as total compensation in industry j, multiplied by the ratio of total compensation in the nonfinancial corporate sector to total compensation in all private industries. We make this adjustment because we want to estimate only the labor rents going to workers in the nonfinancial corporate sector, but we do not have data on compensation by industry broken down by corporate versus noncorporate sector.

firm size wage fixed effect estimates from the CPS for 1990–2016. As with the industry wage premia, to estimate the firm size premium $fsp_{s,t}$ for each firm size class s we halve our estimated firm size (log) wage fixed effects to account for possible compensating differentials or unobserved productivity differences. The firm size premium is estimated for firms of 500+ workers or 100–499 workers, relative to firms with 1–99 workers. We impute firm size rents for the years 1982–1989 using data on compensation share by firm size class and estimated firm size log wage premia from Levine and others (2002).[34] This gives us the following expression for firm size rents:

$$\text{Firm size rents}_t = \sum_{s}^{\text{firm size classes}} \text{compensation}_{s,t}\left(1 - \frac{1}{e^{fsp_{s,t}}}\right)$$

where “compensation” refers to our estimate of nonfinancial corporate sector compensation by firm size class.[35]

Using this method, we think it likely that we will *underestimate* the true decline of labor rents over recent decades. First, because our estimates are based on union, industry, and firm size wage premia calculated relative to a baseline sector (nonunionized firms for union rents, retail trade for industry rents, and firms of under 100 employees for firm size rents), our calculation of total labor rents will miss any decline in rent-sharing that has occurred commonly across industries, firm size classes, or union status. This could include a generalized increase in shareholder activism and more ruthless corporate management practices, a generalized increase in the use of domestic outsourcing, or a generalized decrease in the threat effect of unions. Second, in each calculation we assume that there are no rents in the baseline sector: workers receive the wage that would prevail in the absence of worker power. Our calculation will therefore miss any decline in rent-sharing that is specific to these baseline sectors—with the most obvious candidate being a decline in rents arising from the erosion in the real value of the minimum wage. Third, our estimates of labor rents are based on union, industry, and firm size *earnings* premia. Total rents, however, are estimated as a share of *compensation*. The union and large-firm premia for

34. Full details on the imputation procedure are available in online appendix B.1.

35. This is estimated as total compensation in the nonfinancial corporate sector, multiplied by the payroll share of each firm size class (from the Census Bureau Statistics of US Businesses [SUSB] data).

nonwage benefits are likely greater than for wages, making our calculation of total union and firm size rents an underestimate.[36]

There are, on the other hand, some factors which could make our estimate of the decline in labor rents an overestimate. First, while we cut our estimated industry wage fixed effects and firm size fixed effects in half to account for unobserved productivity or compensating differentials, it is possible that they remain overestimates of the degree of rents (though our benchmarking exercise should assuage this concern). Second, we assume that there are zero rents in the baseline sectors (nonunionized firms, retail trade, and firms of under 100 employees), but in some models, worker power in one sector lowers pay in other sectors (by restricting employment in the high worker power sector, leading workers to spill over into the low worker power sector, reducing wages). If this is the case, we would overestimate total labor rents.[37] On net, we think these concerns are outweighed by the factors pushing our estimate to be an underestimate.

II.D. Labor Rents in the Nonfinancial Corporate Sector, 1982–2016

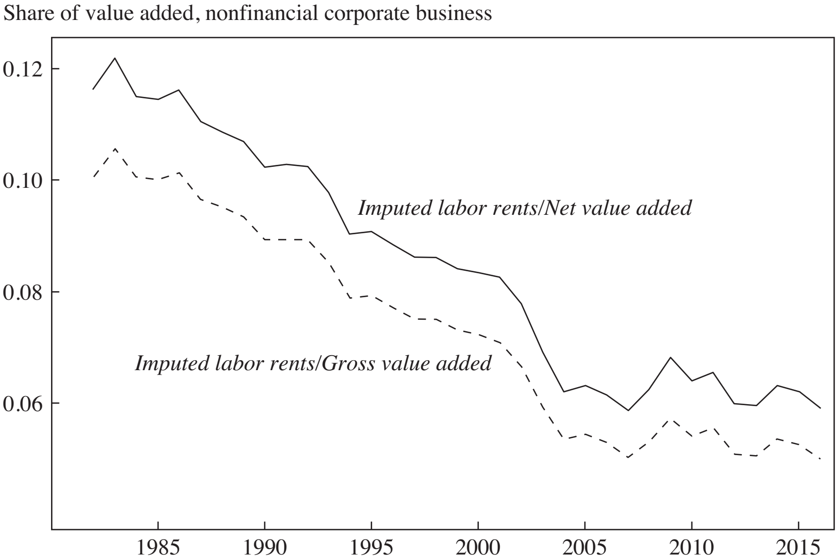
Our measure of labor rents, as a share of net value added in the nonfinancial corporate business sector, declined from around 12 percent in the early 1980s to around 6 percent in the 2010s (figure 5, table 1). Union rents fell by 2.1 percentage points as the unionization rate and union wage premia fell. Industry rents fell by 2.4 percentage points as industry wage premia fell and employment fell in high-rent industries. Firm size rents fell by 1.2 percentage points as firm size premia fell.

A set of simple counterfactuals illustrates that the decline in total labor rents is primarily due to changes in the ability of workers to lay claim to rents *within* any given industry, rather than changes in sectoral composition of the economy. First, if unionization within each sector had *not* fallen (and union wage premia had not fallen), but the sectoral composition of compensation had changed as it did over 1987–2016, union rents would have

36. Mishel and others (2012) show that the union premium is greater for nonwage benefits than for wages. Hollister (2004) finds that large firms are more likely to provide health and pension benefits, controlling on observables, but this differential has fallen over time, exacerbating the fall in the large-firm wage premium.

37. A further concern might be that we estimate union and industry wage effects in the CPS-ORG without controlling for firm size (which is not available in the CPS-ORG). As a robustness check, we estimate union, firm size, and industry wage premia all together in the CPS-ASEC over 1990–2019. The estimated falls in the size of the union wage premium and industry wage premia are very close to those estimated from the CPS-ORG data.

Figure 5. Estimated Labor Rents as Share of Value Added, Nonfinancial Corporate Sector



Sources: BEA NIPA, BEA industry accounts, CPS, SUSB, unionstats.com, authors' calculations.

Table 1. Estimated Labor Rents as Share of Value Added, Nonfinancial Corporate Sector

	1982	*1986*	*1996*	*2006*	*2016*
Shares of net value added (%)					
Total labor rent share	11.7	11.7	8.9	6.2	5.9
Union rent share	3.0	2.6	1.7	1.2	0.9
Firm size rent share	3.7	3.5	2.9	2.3	2.5
Industry rent share	5.0	5.6	4.2	2.6	2.6
Shares of gross value added (%)					
Total labor rent share	10.1	10.2	7.8	5.3	5.0
Union rent share	2.6	2.3	1.5	1.1	0.7
Firm size rent share	3.2	3.0	2.6	2.0	2.1
Industry rent share	4.3	4.9	3.7	2.2	2.2

Sources: BEA NIPA, BEA industry accounts, CPS, SUSB, unionstats.com, authors' calculations.

fallen from 2.4 percent to 1.9 percent over 1987–2016 (rather than falling from 2.4 percent to 0.9 percent).[38] On the other hand, if the sectoral composition of compensation had not changed, but unionization rates within each sector, and union wage premia, had fallen to the levels they were at in 2016, union rents would have fallen by essentially the same amount that they fell in reality: from 2.4 percent to 0.9 percent over 1987–2016.[39] For industry rents, if industry wage premia had not declined but the sectoral composition of compensation had still changed over 1987–2016, the industry rent share would have declined only by around one-tenth of a percentage point.[40] If industry wage premia had fallen but the sectoral composition of compensation had stayed the same, the industry rent share of net value added would have fallen from 5.2 percent in 1987 to 3.4 percent in 2016 rather than from 5.2 percent to 2.6 percent. Finally, for firm size rents, the share of workers in large firms has actually grown over the period, both in aggregate and within almost every sector, such that the decline in firm size rents reflects exclusively the decline in the firm size premium rather than compositional shifts.

Note that our analysis of the role of union rents only considers the *direct* effect of the decline in unionization: the loss of wage premia for unionized workers. To the extent that union power also increased the compensation of nonunion workers in certain industries or large firms through threat effects, our estimates of the decline in industry or firm size rents could also be capturing effects of the decline of unions.[41]

While our analysis in this paper is primarily focused on shifts in income between labor and capital, rather than inequality in labor incomes, we note that the decline in labor rents appears to have disproportionately affected workers with less formal education. Over 1984–2016, labor rents as a share of compensation fell by 8 percentage points for workers with no college or

38. We carry out our counterfactual over 1987–2016 rather than 1982–2016 because it means we are able to use consistently defined NAICS industries. This is the period over which the majority of the fall in labor rents happened.

39. This is because by 2016 the unionization rate in manufacturing had fallen to almost the level that it was in services. So shifting the sectoral composition from services back to manufacturing in 2016 would have made little difference to aggregate unionization.

40. This is due to two offsetting forces. The decline of the share of total compensation in manufacturing—which has a high average wage premium—exerted downward pressure on the industry rent share, but this was offset by increases in the compensation share of professional, scientific, and technical services and by health care and social assistance, which had high and medium-sized wage premia in the late 1980s (respectively).

41. Supporting this, there is a very strong relationship between the decline in industry, firm size, and union rents at the state and industry level. See online appendix C.8 for details.

some college education, and by 5.6 percentage points for workers with a four-year college education or more. This differential was driven by significantly larger declines in unionization rates and firm size rents for non-college-educated workers.[42] There is a large body of work documenting the effect of the decline in unionization on the rise in income inequality in the United States; see, for example, DiNardo, Fortin, and Lemieux (1996), Card (1996), Rosenfeld (2014), Farber and others (2018), and Fortin, Lemieux, and Lloyd (2018).

II.E. Were Labor Rents Redistributed or Destroyed?

One natural explanation for the steep decline of labor rents is that it represents greater market pressures on particular industries, coming from technology, globalization, or some other extrinsic forces. If this were the case, one would expect that returns to capital would fall alongside rents to labor and that the total rents in the industry—profits, plus labor rents—would be falling. It is striking, however, that for the industries in which the majority of the decline in labor rents took place, this was *not* the case—suggesting there was a very important element of redistribution of rents from labor to capital.

In twenty-nine industries—which employed around 30 percent of the private sector workforce in 2018—returns to capital rose even while rents to labor fell over 1987–2016. Together, these industries were responsible for 73 percent of the decline in labor rents over the period. Of these industries, those responsible for the largest shares of the total decline in labor rents were several manufacturing industries, wholesale trade, telecoms, utilities, and trucking. In the majority of these industries—twenty-one industries, employing around 24 percent of the private workforce in 2018—returns to capital rose by more than rents to labor fell over 1987–2016, implying that the total underlying profits generated by these industries rose, even as rents to labor fell. These industries were responsible for 38 percent of the total decline in labor rents over 1987–2016.[43]

We also take a closer look at manufacturing industries. The manufacturing sector can account for the majority of the decline in the labor share since

42. See online appendixes B.3 and C.5 for the detail underlying these calculations. We start in 1984 as we cannot estimate union membership and wage premia by education group before 1984.

43. See online appendix B.6 for details of these calculations. We study fifty-one industries at roughly the NAICS three-digit level, over 1987–2016 (since consistent industry-level data through 2016 are not available before 1987).

the 1980s. It is a sector which saw particularly large declines in unionization and in our estimates of industry wage premia. And it is the sector that has been the most exposed to global competition over recent decades. This raises the question: Were labor rents destroyed most in the manufacturing industries that were most exposed to global competition? Using changes in import penetration from low-wage countries as our measure of exposure to global competition, we investigate this for eighteen manufacturing industries over 1989–2007.[44] Contrary to the predictions of the globalization thesis, labor rents declined the most in the industries with the smallest increases in low-wage import penetration over the period (figure 6). This evidence, while not dispositive, casts further doubt on the argument that the decline in labor rents in manufacturing since the late 1980s was primarily a result of globalization.

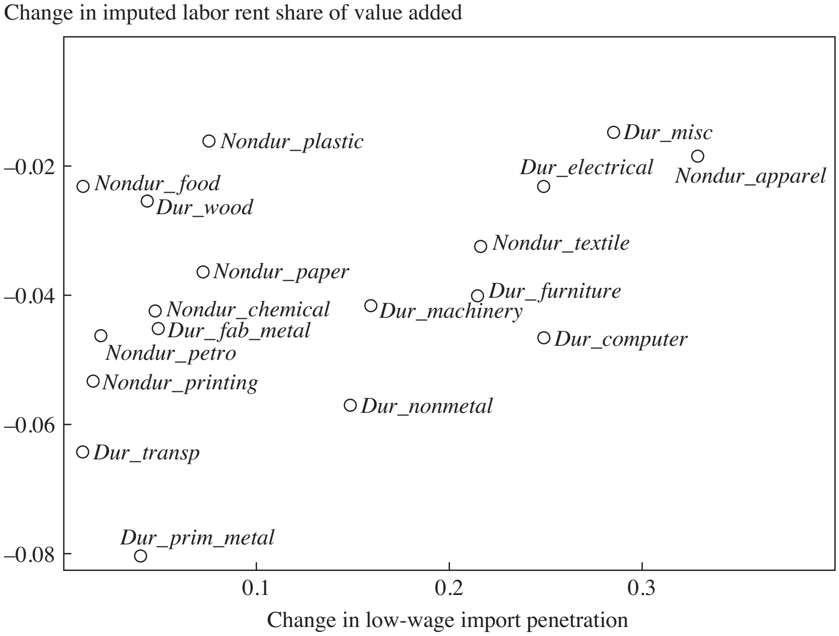
Overall, these results suggest that a large share of the decline in labor rents was a result of a redistribution of rents from labor to capital, rather than a destruction of rents as a result of increased competition or market pressure. This informs our approach in the rest of the paper.

III. Factor Shares, Profits, and Measured Markups

The labor share of income has declined since the 1980s, with a corresponding rise in the capital share (Elsby, Hobijn, and Şahin 2013; Karabarbounis and Neiman 2014). The Tobin's q of publicly listed corporations—the ratio of their stock market value to the replacement cost of their capital

44. We use low-wage import penetration data from Bernard, Jensen, and Schott (2006), updated by Peter Schott in 2011. Low-wage import penetration is calculated as the share of domestic sales within each industry represented by imports from low-wage countries, defined as countries with GDP per capita less than 5 percent of the US level. We study 1989–2007 as this is the period for which we have consistently defined data on low-wage import penetration (see online appendix B.7 for more details). Our sample period covers the period after the accession of China into the WTO, as well as the large increases in global trade in the 1990s. However, our sample period does not cover the effects of globalization in the 1970s and early to mid-1980s. Competition from low-wage countries would have been relevant for only a few industries during this period: in 1989, imports from low-wage countries only made up more than 1 percent of the US market in three manufacturing industries: apparel, textiles, and miscellaneous durable goods (Bernard, Jensen, and Schott 2006). On the other hand, competition from high-wage countries may have destroyed rents in other manufacturing industries earlier in the postwar period, and this is not captured in our sample. Borjas and Ramey (1995), for example, argue that increased foreign competition in durable goods manufacturing over 1976–1990 destroyed rents in that sector, reducing the wage premia paid to workers.

Figure 6. Low-Wage Import Penetration and Labor Rents in Manufacturing, 1989–2007



Sources: CPS-ORG; BEA; Bernard, Jensen, and Schott (2006); authors' calculations.
Note: Manufacturing industries at BEA industry code level.

stock—has risen from around 1.0 in 1970 to 1.75 by 2015, alongside an increase in the value of financial assets relative to the value of productive capital (Eggertsson, Robbins, and Wold 2018). The average profitability of capital has risen, even as the risk-free rate has declined. And, by a range of measures, several authors have found that markups have risen (De Loecker, Eckhout, and Unger 2020; Eggertsson, Robbins, and Wold 2018; Covarrubias, Gutiérrez, and Philippon 2019).[45]

A number of explanations have been proposed for the decline in the labor share of income. Many of these have centered on certain aspects of globalization or technological change—such as the increase in offshoring, the declining price of capital goods, or rising automation—as the major

45. The magnitude of the rise in measured markups depends on the method used. See Traina (2018), Karabarbounis and Neiman (2018), Edmond, Midrigan, and Xu (2018), and Baqaee and Farhi (2020). All measures that we are aware of show some increase in markups over recent decades.

cause of the decline in the labor share. Papers that focus on the United States include Elsby, Hobijn, and Şahin (2013), Abdih and Danninger (2017), Acemoglu and Restrepo (2018), and Autor and others (2020); those taking a cross-country perspective include Karabarbounis and Neiman (2014), Dao and others (2017), and Autor and Salomons (2018).

More recently, a growing body of research argues that these trends can be explained by a rise in the market power of corporations. Rising monopoly power in product markets would lead firms to increase their markups, reducing the labor share of income and increasing corporate profitability. This would in turn increase Tobin's q and the value of financial assets relative to physical capital. Different aspects of this argument have been made by Barkai (forthcoming), Brun and González (2017), Covarrubias, Gutiérrez, and Philippon (2019), De Loecker and Eeckhout (2019), De Loecker, Eeckhout, and Unger (2020), Eggertsson, Robbins, and Wold (2018), Farhi and Gourio (2018), González and Trivín (2019), Grullon, Larkin, and Michaely (2019), Gutiérrez and Philippon (2017, 2019), Hall (2018), and Philippon (2020). Some authors have also argued that these trends could be rationalized by a rise in companies' monopsony power in labor markets (CEA 2016; Furman and Krueger 2016; Glover and Short 2018; Benmelech, Bergman, and Kim 2018; Philippon 2020).

It is difficult to rationalize the trends in corporate valuations, corporate profitability, and measured markups in a model of perfect competition. In this sense, we agree with the monopoly/monopsony power arguments that the explanation of these macro trends must involve some degree of rents created by imperfect competition (in contrast to explanation based solely on technological change or globalization).

Our preferred explanation for these macro trends, however, focuses on a *redistribution* of existing rents rather than a *creation* of new rents. That is, the decline in the labor share—and the rise in corporate valuations, profitability, and measured markups—could have been caused by a decline in worker power.

To see this, consider an economy characterized by three types of power, to varying degrees: monopoly power, monopsony power, and worker power.

Firms have monopoly power in the product market, created by a combination of monopolistic competition and restrictions to entry. They set their price at a markup above marginal cost and make some pure profits, or rents, which are not fully competed away by new entrants. These rents may arise as a result of explicit barriers to entry, regulatory or otherwise. But they may also arise from heterogeneous production technologies, with new entrants unable to perfectly replicate incumbents' products or production techniques. And in the short run, there may be rents because of the presence of fixed costs due to previously installed capital and prices in excess of variable costs.[46]

Firms may also have monopsony power in the labor market, by which we mean the wage-setting power firms derive from an upward-sloping labor supply curve. This can arise either from employers' size in their local labor market (conventional monopsony) or from labor market search frictions, switching costs, or different worker preferences for different employers (dynamic monopsony). In a monopsonistically competitive labor market, the wage a firm pays is a markdown from the marginal revenue product of labor at the firm.[47]

Finally, there is also worker power. By worker power, we mean workers' ability to increase their pay above the level that would prevail in the absence of such bargaining power. In this framework, worker power not only acts as countervailing power to firm monopsony power but also gives workers an ability to receive a share of the rents generated by companies operating in imperfectly competitive product markets. We use the term *worker power* as synonymous with worker bargaining power, worker rent-sharing power, and insider-outsider power of the kind that was used in earlier work to explain increases in unemployment.[48]

In this framework, if workers' ability to receive some of the rents generated by their firms has fallen over time, we would expect to see a decline

46. Note that in the latter two cases the existence of rents does not necessarily signal a market imperfection that can be corrected through antitrust or competition policy. In this framework the presence of rents is therefore, to some extent, an innate feature of the structure of particular product markets.

47. Our definition of monopsony power follows the modern monopsony literature. In the presence of monopsony, the size of the wage markdown is an inverse function of the elasticity of labor supply to the firm. The perfectly competitive case occurs where the elasticity of labor supply to the firm is infinite. Labor market concentration and search frictions both therefore create monopsony power because they both generate upward-sloping labor supply curves to the firm—but their welfare and policy implications can be different, as highlighted by Manning (2003).

48. Note that monopsony power and worker power are distinct concepts in our framework. The term *monopsony power* is sometimes used to refer to a broader conception of employer power than we use here; for example, in some bargaining models, firm monopsony power might be considered the exact inverse of worker power (the wage is partly determined by the firm's and worker's relative bargaining power over the match surplus). We distinguish between monopsony power and worker power for two reasons. First, in our framework, worker power is not necessarily simply the inverse of employer wage-setting power; worker power enables workers to claim a share of the rents produced within the firm, potentially raising their wage above the marginal product in their labor market. This can occur even in a world of no labor market concentration or search frictions, where labor supply to the firm is completely elastic. Second, the source of the change in wage-setting power matters for

in the labor share—as rents going to workers fall and rents going to shareholders rise (holding constant the total quantity of rents generated). We would also expect to see a divergence between the average profitability of capital and the risk-free rate, as profits to shareholders rise, and a rise in Tobin's q and the ratio of financial wealth to physical capital, as the rise in profits to shareholders increases the net present value of the claim shareholders have over corporate profits (even as the asset value of firms does not change). Indeed, Greenwald, Lettau, and Ludvigson (2019) find that a reallocation of income from labor to shareholders can account for a large share of the rise in equity valuations from 1989 to the present.[49]

In addition, while a fall in worker rent-sharing power should not have any implication for firms' underlying markups (which are determined by their product market power), it does have implications for *measured* markups. This is because measures of aggregate markups used in recent literature depend on firms' costs, including firms' labor costs—even if the labor costs partly represent rents accruing to labor as well as the true marginal cost of production.[50] This implies that markups, as they have

diagnosis and policy solutions; a decline in worker power caused by a decline in unionization implies a different policy solution as compared to a rise in employer power caused by an increase in labor market frictions or concentration. The two concepts of worker power and monopsony power are, however, linked in the sense that worker power operates as countervailing power to firm monopsony power. As worker power declines, firms' ability to exercise their monopsony power rises without the underlying elasticity of labor supply to the firm having changed, as described in, for example, Erickson and Mitchell (2007).

49. Specifically, they find that a series of "factor share shocks" have reallocated rewards to shareholders and away from labor compensation, accounting for 43 percent of the increase in equity valuations since 1989. They do not take a stance on the cause of these factor share shocks but note that they could be due to changes in industrial concentration, worker bargaining power, offshoring and outsourcing, or technological change.

50. The production function approach used by De Loecker, Eeckhout, and Unger (2020) estimates markups as a function of the (estimated) elasticity of output with respect to variable inputs and the ratio of sales to variable costs—which include some labor costs. The rise in measured markups in the United States is mostly due to an increasing ratio of sales to variable costs, which could be a result of falling labor costs as labor rents fell. The user cost approach of Gutiérrez and Philippon (2017) estimates markups as the ratio of sales to costs, which are calculated as operating expenses plus an imputed cost of capital. Operating expenses include labor costs. Again, this means that changes in measured markups could be due to changes in labor costs as a result of falling labor rents (see online appendix F for more details on this). It would in theory be feasible to take these approaches and apply them only to nonlabor costs to estimate markups, but there are no publicly available data of sufficiently good quality to do this across the entire set of industries over a long time period. One example of this approach is Anderson, Rebelo, and Wong (2019) who study the retail trade industry, estimating the markups of the price of each good sold over its replacement cost, *not* including labor costs.

been measured in recent papers, cannot be used to distinguish between a story of rising product market power and a story of falling worker power: a rise in measured markups could reflect a fall in worker rent-sharing power just as much as it could reflect a rise in true markups and firms' monopoly power.

III.A. Accounting Decomposition

This implies that rising monopoly power, rising monopsony power, and falling worker power could each in theory account for the changes in factor shares, profits, and markups. But is the magnitude of the decline in labor rents consistent with these trends? To calibrate the plausibility of the declining labor rents explanation, we build on the accounting decomposition in Farhi and Gourio (2018). Farhi and Gourio extend the neoclassical growth model to account for six major recent macroeconomic trends, including the decline in the labor share, increases in valuation ratios, and moderate increases in profitability alongside a declining risk-free rate. Using this model, they identify a role for rising monopoly power in explaining these macro trends (alongside roles for unmeasured intangibles and rising risk premia). They estimate that average economy-wide markups rose from 8 percent to 15 percent over 1984–2016.

Their model, however, assumes competitive labor markets with no rent-sharing. We replicate their accounting decomposition, with one alteration: we hold the degree of monopoly power (markups) fixed and instead introduce a rent-sharing parameter to allow workers to share in monopoly profits. We incorporate this in the simplest way possible: the monopolistic representative firm maximizes profits as before but then shares the rents, or pure profits, with share π_L going to labor. This reduced-form approach is similar to that adopted in much of the literature on rent-sharing, as reviewed in Card and others (2018). It can be microfounded with a strongly efficient bargaining model where workers, seeking to maximize total pay to labor, and shareholders, seeking to maximize their profits, jointly bargain over the firm's production decisions (MacDonald and Solow 1981).

Farhi and Gourio (2018) carry out their decomposition targeting nine empirical moments for the US private sector over 1984–2016: gross profitability, the gross capital share, the investment-capital ratio, the risk-free rate, the price-dividend ratio, population growth, total factor productivity (TFP) growth, the growth rate of investment prices, and the employment-population ratio. They estimate nine parameters: the discount factor, the probability of a disaster, the depreciation rate of capital, the Cobb-Douglas parameter in the aggregate production function, population growth rate, TFP growth, the growth rate of investment-specific productivity, labor supply, and the markup.

We target the same nine moments and estimate eight of the same nine parameters—but, instead of estimating the markup, we estimate the rent-sharing parameter with labor, holding the markup fixed at the level that Farhi and Gourio estimate for the period 2001–2016 (1.15). Identification is nearly recursive in the Farhi and Gourio decomposition, with many parameters estimated tightly by their near-equivalent moments. Identification in our approach is therefore nearly identical to that in Farhi and Gourio: it has different implications for only two of the nine empirical moments—the gross capital share $\frac{\Pi}{Y}$ and gross profitability $\frac{\Pi}{K}$ (equivalent in the Farhi and Gourio model to the marginal product of capital). The equations below show the difference between the two approaches: in the Farhi and Gourio model, the rent-sharing parameter π_L is implicitly set to be constant at zero, and the markup μ is allowed to vary. In contrast in our model, the markup μ is set to be constant at 1.15, and π_L is allowed to vary.

$$\text{Capital share } \frac{\Pi}{Y} = \frac{\alpha + (1 - \pi_L)(\mu - 1)}{\mu}$$

$$\text{Profitability of capital } \frac{\Pi}{K} = \frac{\alpha + (1 - \pi_L)(\mu - 1)}{\alpha}\left(r^* + \delta + g_Q\right)$$

By construction of the recursive identification process in the decomposition, our model returns exactly the same parameter estimates as Farhi and Gourio for six of the nine parameters estimated. Table 2 shows only the parameter estimates that differ between the Farhi and Gourio model ("FG") and our model ("SS"). To fit the data best, Farhi and Gourio estimate a rise in the average economy-wide markup from 1.08 to 1.15 over the period. When we hold the markup constant at 1.15, but allow the rent-sharing parameter to vary, we estimate instead that the rent-sharing parameter fell from 0.44 to 0.02 over the period.[51] Our model also has slightly different implications for the Cobb-Douglas parameter α and TFP growth g_Z: our model suggests a somewhat smaller slowdown in TFP growth over the period, and a slight fall in the Cobb-Douglas parameter α (implying a small degree of labor-complementing technological change).

51. A rent-sharing parameter of 0.44 is quite plausible when compared to the range of estimates from studies of rent-sharing. See online appendix D.2 for details.

Table 2. Estimated Parameters and Changes over Time

Parameter		*Model*	*First sample (1984–2000)*	*Second sample (2001–2016)*	*Difference*
Markup	μ	FG	1.079	1.146	0.067
		SS	1.15	1.15	—
Rent-sharing with labor	π_L	FG	0	0	—
		SS	0.441	0.022	–0.419
Cobb-Douglas parameter	α	FG	0.244	0.243	–0.001
		SS	0.260	0.244	–0.016
TFP growth	g_Z	FG	1.298	1.012	–0.286
		SS	1.233	1.010	–0.223

Source: Farhi and Gourio (2018); authors' calculations.

Note: Only parameters where our estimates differ from Farhi and Gourio's (2018) estimates are shown. In the "SS" estimation, markup μ is held constant at 1.15. In the "FG" estimation, the rent-sharing parameter π_L is implicitly held constant at zero. The "FG" estimates in this table correspond to the baseline parameter estimates in table 2 of Farhi and Gourio (2018).

What does the estimated fall in the rent-sharing parameter imply for total labor rents? The rise in markups estimated by Farhi and Gourio, from 1.08 to 1.15, implies a rise in the pure profit share of output from 7.3 percent in the 1980s and '90s to 12.8 percent in the 2000s and 2010s. Since we hold the markup at 1.15 through these four decades in our estimation, the pure profit share of our economy is, by construction, 12.8 percent throughout 1982–2016. The estimated fall in the rent-sharing parameter therefore implies that the share of gross private sector output that was labor rents fell by 5.3 percentage points, from 5.6 percent to 0.3 percent, over the period. This is quite similar to our estimate of the decline of labor rents in section II; we estimated that labor rents fell by 5.1 percentage points of gross value added in the nonfinancial corporate sector over 1982–2016 (corresponding to a fall of 4.1 percentage points of gross business sector value added).[52] There is no necessary reason why these two estimates should line up so closely: the estimate of the fall in labor rents from the Farhi and Gourio model comes from the best fit of nine parameters to nine macro moments in each of the two periods, while our estimate of the fall in labor rents comes from our estimated union, industry, and firm size wage premia using CPS data. (Note that to match Farhi and Gourio's results we set up

52. Nonfinancial corporate sector value added was 72 percent of total business sector value added in 1982 and 65 percent in 2016 (BEA NIPA tables 1.3.5 and 1.14). This calculation assumes that the only change in labor rents occurred in the nonfinancial corporate sector, that is, that there was no change in labor rents elsewhere in the business sector.

our calibration such that labor rents must equal zero in the second period. Therefore, the percentage point change in the share of output represented by labor rents is a more appropriate comparator than the levels.)

We see this accounting exercise as suggesting that (1) the degree of the fall in rent-sharing with labor which is required to be consistent with a number of key macro moments over 1982–2016 is both relatively consistent with our empirical estimates of the actual fall in rent-sharing with labor, and relatively consistent with estimates of rent-sharing elasticities from the micro literature; and (2) despite the differential implications for investment of a rise in monopoly power versus a fall in rent-sharing, when incorporated into a full general equilibrium model it is possible to reconcile a fall in labor rent-sharing (in an efficient-bargain type framework) with the data on capital and investment, without implausible implications for other macro variables.

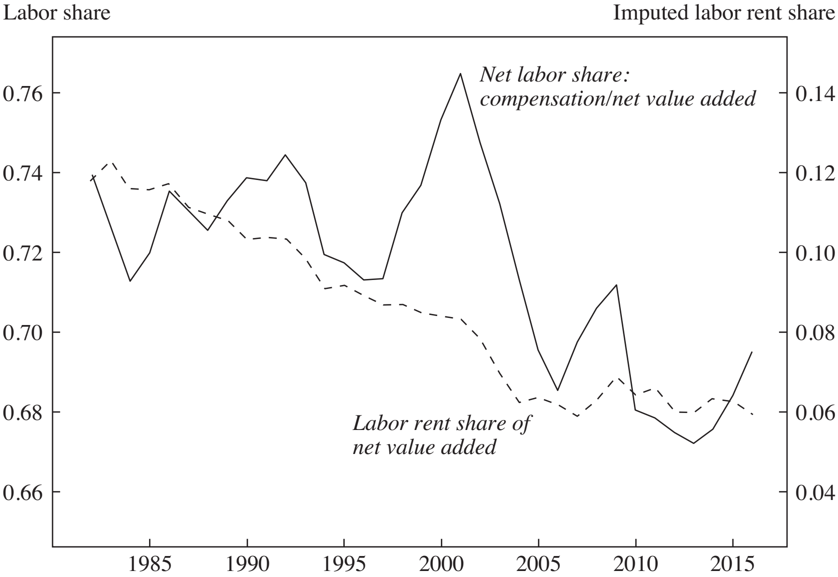
III.B. Aggregate and State-Level Evidence: Factor Shares

Next, we compare our estimates of the decline in the labor rent share of value added with aggregate changes in factor shares. The net labor share in the nonfinancial corporate sector (compensation over net value added) fell by 4.4 percentage points over 1982–2016.[53] Our measure of the labor rent share of net value added in the nonfinancial corporate sector fell by almost 6 percentage points over the same period. This suggests that the decline in imputed labor rents as estimated from industry, union, and firm size wage premia can more than fully explain the decline in the net labor share over the period (as shown in figure 7); that is, the entirety of the shift in the functional income distribution in the nonfinancial corporate sector could be explained by a redistribution of rents from labor to capital.

The other side of the coin of the fall in the labor share is the rise in the capital share. Since our measure of labor rents can be interpreted as a measure of the firm's profits which go to labor, with the rest of the firm's profits going to capital, we can define the total profit share of value added as the share of value added accounted for by capital income *plus* labor rents. While the capital share of net value added has risen over 1982–2016, our imputed measure of the total profit share has stayed roughly constant or even fallen slightly (figure 8)—consistent with the interpretation that the

53. Following Bridgman (2018) and others, for our main results at the aggregate and industry levels we use the labor share of value added net of depreciation, as the depreciation rate has risen over the period.

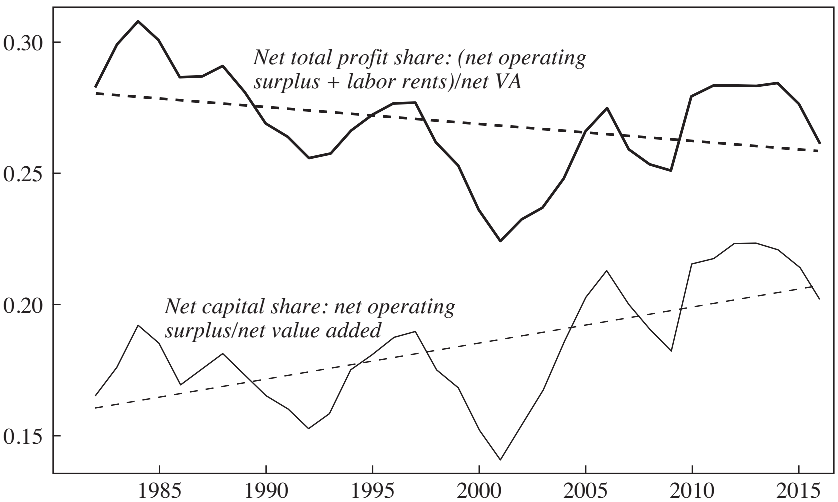
Figure 7. Net Labor Share and Imputed Labor Rent Share, Nonfinancial Corporate



Sources: BEA NIPA; authors' calculations.

Note: Labor share refers to the compensation share of net value added in the nonfinancial corporate sector. Our measure of the imputed labor rent share of net value added is calculated as described in section II.

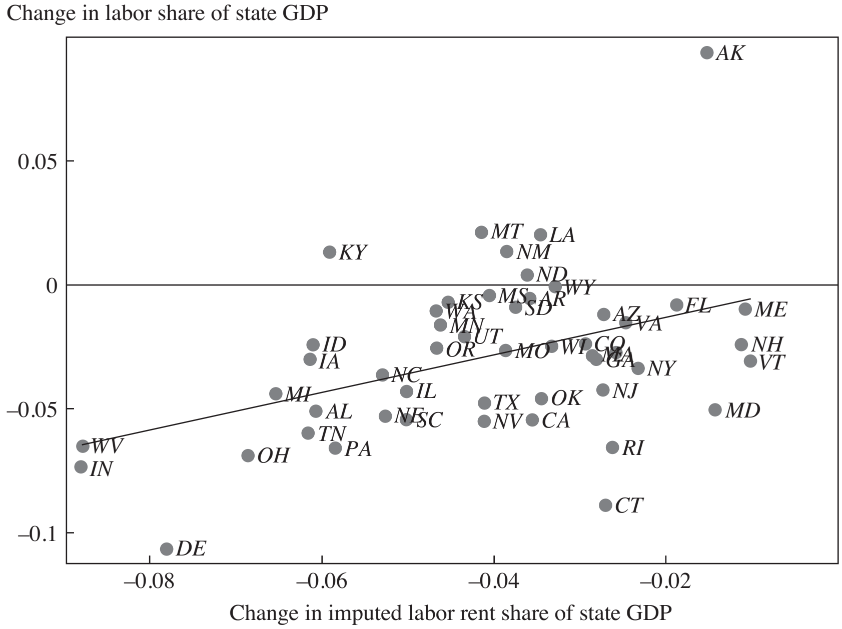
Figure 8. Net Capital Share and Imputed Profit Share, Nonfinancial Corporate



Sources: BEA NIPA; authors' calculations.

Note: Net capital share in the nonfinancial corporate sector is calculated as net operating surplus over net value added. Measure of the net total profit share is calculated as the net operating surplus plus our measure of imputed labor rents (explained in section II), divided by net value added. Dashed lines are lines of best fit.

Figure 9. Changes in State-Level Labor Share and Labor Rent Share, 1984–1988 to 2012–2016



Sources: BEA regional economic accounts; authors' calculations.

Note: Imputed labor rent share of state GDP is calculated from estimated union, firm size, and industry wage premia; state-level unionization rates; and compensation by industry. Labor share of state GDP is defined as state-level compensation over GDP.

total profitability of firms (and their monopoly power) has not risen over the period but that these profits have instead partly been redistributed from labor to capital.

We observe a similar pattern with state-level data. Estimating state-level labor rent shares in the same way as we estimate the aggregate labor rent share, we show that states with bigger declines in their imputed labor rent share also saw bigger declines in their labor share over 1984–2016 (figure 9).[54] This strong relationship persists in regressions at the annual level,

54. The coefficient in a regression of the change in the state labor share over 1984–1988 to 2012–2016 on the change in the labor rent share over the same period is 0.76, with a p value of 0.002 and an R^2 of 0.19. We calculate the labor share as state-level compensation over GDP and calculate labor rents as a share of state GDP, using data from the BEA Regional Economic Accounts. We start in 1984 because it is the first year for which we can estimate state-level unionization and union wage premia. More details are in online appendix B.4.

Table 3. State-Level Regressions of Labor Share on Measures of Labor Power

Regression of labor share of state GDP on imputed labor rent share of state GDP, 1984–2016				
Imputed labor rent share	0.94**	1.09**	0.69**	0.52**
	(0.14)	(0.28)	(0.06)	(0.13)
Fixed effects	None	Year	State	Year, State
Observations	1,650	1,650	1,650	1,650
State-level regression of labor share on imputed union rent share of state GDP, 1984–2016				
Imputed union rent share	1.76**	1.46*	1.98**	1.04*
	(0.48)	(0.68)	(0.24)	(0.40)
Fixed effects	None	Year	State	Year, State
Observations	1,650	1,650	1,650	1,650

Source: Authors' calculations.
Note: Robust standard errors, clustered at state level, in parentheses.
*$p < 0.05$, **$p < 0.01$

with year and state fixed effects, as shown in table 3 (both for the labor rent share and for the union rent share component of it).[55]

III.C. Industry-Level Evidence

Next, we estimate labor rents at the level of fifty-one industries over 1987–2016.[56] We analyze the relationship between industry-level changes in labor rents and changes in the labor share, profitability, and Tobin's q. Since a number of recent papers have highlighted the link between industrial concentration and changes in labor shares and profitability, we also incorporate product market concentration, using measures of industry-level top twenty import-adjusted sales concentration calculated from Compustat and census data by Covarrubias, Gutiérrez, and Philippon (2019).[57]

55. Similarly, Hazell (2019) finds that right-to-work laws (which reduce union power) reduce state-level labor shares.

56. Our industry definitions are very close to the BEA industry codes (roughly NAICS three-digit; see online appendix G for more details on industry definitions). For consistency with the previous section, we do not analyze industries in finance, insurance, and real estate. We also follow Covarrubias, Gutiérrez, and Philippon (2019) in omitting the industry "management of companies and enterprises." Our calculation of industry rents and union rents follows the description in section II closely, with the exception that it is comprised only of union rents and industry rents (and not firm size rents), as we do not have data on compensation shares by firm size class and industry (see online appendix B.5 for more details). Note that for industry rents, the wage premium is estimated relative to the lowest-wage large industry, which is food services and drinking places.

57. We are grateful to Germán Gutiérrez and Thomas Philippon for sharing with us the measures of concentration they constructed for Covarrubias, Gutiérrez, and Philippon (2019). They construct the top four, eight, twenty, and fifty import-adjusted sales concentration ratios

Our analysis shows that over 1987–2016 industries with larger falls in their imputed labor rent share also saw substantially larger falls in their labor share (figure 10).[58] There is a negative, though somewhat weaker, relationship between changes in the labor share and average top twenty import-adjusted sales concentration (figure 11).

We regress the gross and net labor share on the imputed labor rent share of industry value added and on product market concentration at the annual level over 1987–2016, including different combinations of year and industry fixed effects (table 4). Coefficients on the labor rent share are large, positive, and highly significant, and coefficients on concentration are negative and mostly significant.

What is the explanatory power of the decline in labor rents relative to the rise in concentration? Over 1997–2012 (the period for which we have the more accurate census-based concentration data, and in which Covarrubias, Gutiérrez, and Philippon (2019) argue concentration has led to rising monopoly power) the average industry saw a fall in its labor share of 5.2 percentage points. Using the coefficient from the specification with industry and year fixed effects, the average industry's fall in their labor rent share over 1997–2012 was associated with 4.3 percentage points fall in the labor share. The average industry-level increase in import-adjusted top twenty sales concentration was associated with a 0.5 percentage point fall in the labor share.[59] This suggests that declining labor rents can explain the

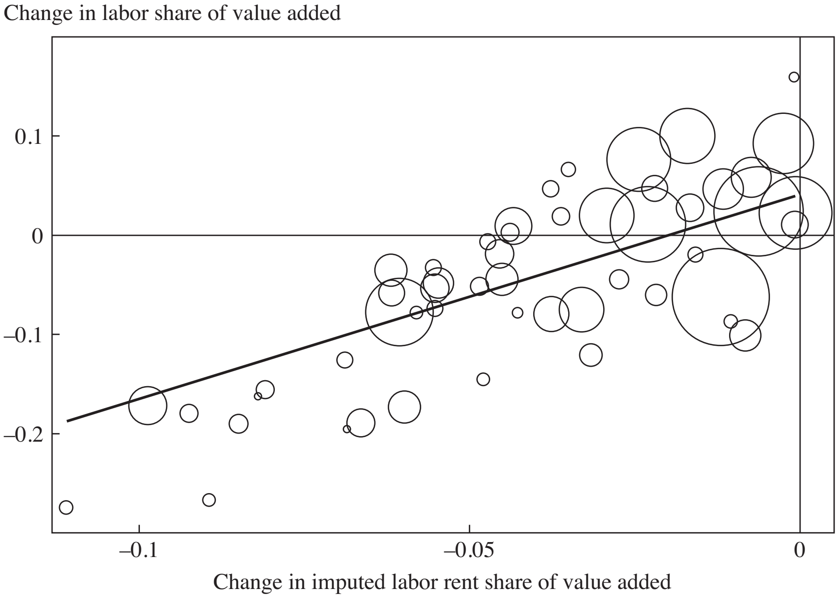
---

for each of the fifty-three BEA industries. They use two data sources: Compustat data on publicly listed companies, reweighted to reflect the composition of the underlying economy, and census data on all firms. The Compustat concentration ratios are available annually for our whole sample period (1987–2016). The census concentration ratios are available for the years 1997, 2002, 2007, and 2012. They adjust for imports by multiplying the domestic sales concentration ratio by the share of US-produced goods in total domestic sales in that industry. More details on the construction of these variables are available in Covarrubias, Gutiérrez, and Philippon (2019). Note that the Compustat measure only covers publicly traded firms, and trends in publicly traded firms have not always been representative of aggregate trends within individual industries (Davis and others 2006). Concentration is an imperfect measure of firms' market power (Berry, Gaynor, and Scott Morton 2019; Syverson 2019). We use concentration in this paper because recent literature has noted the rise in concentration, alongside rising markups and falling labor shares, and has often interpreted this as rising monopoly power.

58. A similar relationship exists for changes in the industry-level unionization rate. See online appendix C.7.

59. The average industry's labor rent share declined by 1.6 percentage points over 1997–2012. Multiplying this by 2.67 suggests a decline in the labor share of 4.3 percentage points. The average industry's import-adjusted top twenty sales concentration rose by 1.8 percentage points over 1997–2012. Multiplying this by −0.28 suggests a decline in the labor share of 0.5 percentage points.

Figure 10. Change in Labor Share and Imputed Labor Rent Share, by Industry, 1988–1992 to 2012–2016



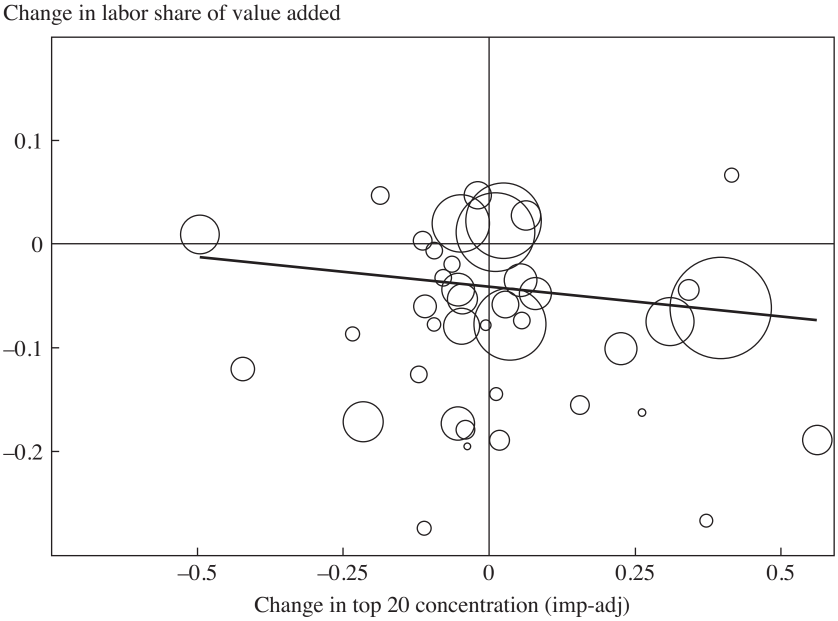
Sources: BEA industry accounts; authors' calculations.
Note: Each bubble is an industry (at BEA industry code level). Bubble size represents industry average employment over 2012–2016. The solid line is an employment-weighted line of best fit.

majority of the average fall in labor shares at the industry level, whereas the average increase in concentration can explain only around 10 percent.[60]

Next, we analyze our measures of labor power alongside three measures of profitability at the industry level over 1987–2016: the gross profit rate (defined as gross operating surplus over fixed assets), as well as two measures of Tobin's q calculated from firm-level Compustat data by Covarrubias, Gutiérrez, and Philippon (2019): the weighted average Tobin's q across publicly listed firms within an industry ("aggregate q"),

60. The relative explanatory power of the worker power measures versus concentration measures is similar if we use other measures of concentration (top four, eight, or fifty sales ratios, and using measures from the census versus Compustat). The comparison of coefficient magnitudes is even starker over 1987–2016: the average fall in the labor rent share was associated with a 10.1 percentage points fall in the labor share, while the average increase in import-adjusted top twenty sales concentration over this period was associated with a 0.1 percentage point fall in the labor share. The average industry's fall in the labor share over this period was 5.2 percentage points.

Figure 11. Change in Labor Share and Top Twenty Sales Concentration (Import-Adjusted), by Industry, 1988–1992 to 2012–2016



Sources: Labor share computed from BEA industry accounts; concentration calculated from Compustat data by Covarrubias, Gutiérrez, and Philippon (2019).

Note: Each bubble is an industry (at BEA industry code level). Bubble size represents industry average employment over 2012–2016. The solid line is an employment-weighted line of best fit.

and the median firm q.[61] Figures 12 and 13 illustrate that over the whole period, falling labor rent shares were associated with rising gross profitability, while rising concentration was associated with rising profitability. In horse-race regressions of profitability measures on our measures of imputed labor rents and industrial concentration (table 4), coefficients on the imputed labor rent share are almost all negative and, for the 1987–2016 regressions, mostly statistically significant.[62] Coefficients on the concentration measures, on the other hand, are mostly not significant and often

61. Results are very similar when we use the simple average Tobin's q across firms, rather than the weighted average.

62. This is consistent with Salinger (1984), who argued that in the 1980s, Tobin's q was low in industries with high monopoly power because unionized workers received the monopoly rents.

Table 4. Industry-Level Regressions—Labor Shares, Profitability, and Investment-to-Profits

Regressions of labor shares and investment-profit on labor rent share and Compustat concentration. N = *1,189 (forty-one industries, 1987–2016)*

Dependent variable	*Labor share of gross value added*				*Labor share of net value added*				*Investment-to-profit ratio*			
Imputed labor rent share of gross value added[a]	2.24** (0.50)	2.44** (0.64)	1.81** (0.16)	2.22** (0.26)	2.40** (0.39)	2.62** (0.47)	2.56** (0.32)	3.65** (0.44)	3.12** (1.12)	3.78* (1.47)	2.26** (0.68)	5.74** (1.30)
Average top twenty sales concentration imp-adj. (Compustat)	–0.23** (0.06)	–0.23** (0.07)	–0.05 (0.05)	–0.05 (0.05)	–0.19* (0.08)	–0.21* (0.08)	–0.06 (0.09)	–0.07 (0.07)	0.25 (0.19)	0.24 (0.20)	–0.15 (0.22)	–0.16 (0.20)
Fixed effects	None	Year	Ind	Yr, Ind	None	Year	Ind	Yr, Ind	None	Year	Ind	Yr, Ind

Regressions of profitability on labor rent share and Compustat concentration. N = *1,189 (forty-one industries, 1987–2016)*

Dependent variable	*Gross profit rate*				*Aggregate Q*				*Median Q*			
Imputed labor rent share of gross value added	–0.80 (0.55)	–1.03 (0.71)	–0.60* (0.28)	–1.93** (0.55)	–3.66** (1.14)	–3.16* (1.33)	–3.17** (0.93)	1.58 (1.51)	–3.22** (0.99)	–2.46* (1.16)	–4.43** (0.86)	–1.24 (1.36)
Average top twenty sales concentration imp-adj. (Compustat)	–0.11 (0.10)	–0.10 (0.10)	0.02 (0.13)	0.04 (0.12)	0.02 (0.12)	0.01 (0.13)	–0.31 (0.32)	–0.31 (0.30)	0.19 (0.13)	0.18 (0.13)	0.15 (0.20)	0.16 (0.20)
Fixed effects	None	Year	Ind	Yr, Ind	None	Year	Ind	Yr, Ind	None	Year	Ind	Yr, Ind

Regressions of labor shares and investment-profit on labor rent share and census concentration. N = *174 (forty-five industries for 1997, 2002, 2007, 2012)*

Dependent variable	*Labor share of gross value added*				*Labor share of net value added*				*Investment-to-profit ratio*			
Imputed labor rent share of gross value added[a]	1.88** (0.65)	1.95** (0.72)	2.18** (0.37)	2.67** (0.42)	2.45** (0.49)	2.56** (0.54)	3.14** (0.43)	3.76** (0.48)	4.32* (1.76)	4.52* (1.91)	5.29⁺ (2.89)	7.24⁺ (4.29)
Average top twenty sales concentration imp-adj. (census)	–0.51** (0.09)	–0.52** (0.09)	–0.24* (0.11)	–0.28* (0.11)	–0.45** (0.12)	–0.46** (0.12)	–0.40* (0.18)	–0.45** (0.16)	0.49 (0.42)	0.48 (0.42)	–0.85 (0.68)	–0.88 (0.78)
Fixed effects	None	Year	Ind	Yr, Ind	None	Year	Ind	Yr, Ind	None	Year	Ind	Yr, Ind

Regressions of profitability measures on labor rent share and census concentration. N = *174 (forty-five industries for 1997, 2002, 2007, 2012)*

Dependent variable	*Gross profit rate*				*Aggregate Q*				*Median Q*			
Imputed labor rent share of gross value added	–0.75 (0.81)	–0.84 (0.85)	–1.35* (0.58)	–2.84** (0.55)	–1.79 (1.52)	–2.15 (1.62)	0.52 (2.20)	–2.22 (3.06)	–0.79 (1.16)	–0.22 (1.14)	–3.25* (1.48)	–1.56 (1.95)
Average top twenty sales concentration imp-adj. (census)	–0.45 (0.33)	–0.44 (0.33)	0.35 (0.29)	0.45 (0.31)	–0.47* (0.22)	–0.43⁺ (0.22)	–1.49⁺ (0.86)	–0.89 (0.71)	–0.09 (0.16)	–0.10 (0.16)	–0.91⁺ (0.47)	–0.39 (0.42)
Fixed effects	None	Year	Ind	Yr, Ind	None	Year	Ind	Yr, Ind	None	Year	Ind	Yr, Ind

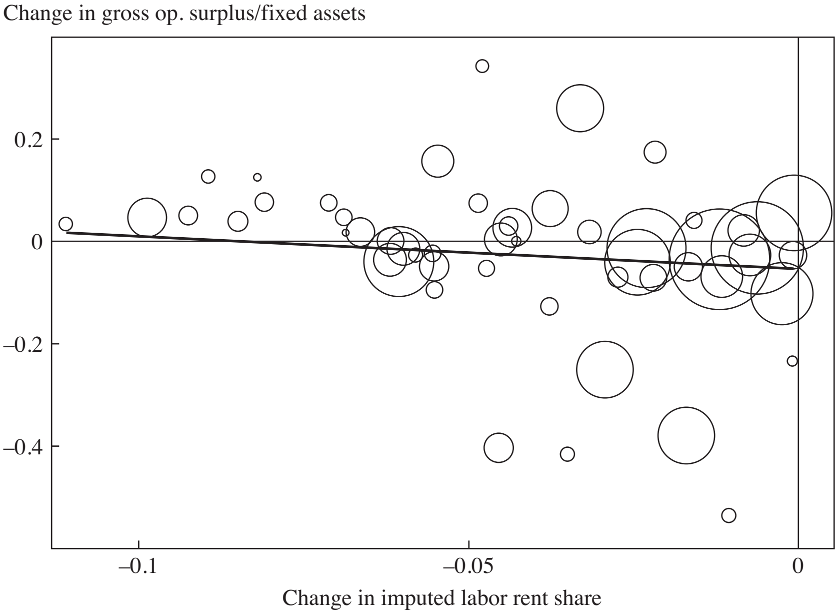
Source: Authors' calculations.

Note: Robust standard errors, clustered at industry level, in parentheses. Investment-profits are 98 percent Winsorized. Regressions are for forty-one or forty-five industries because we do not have concentration data for all fifty-one nonfinancial industries.

a. Imputed labor rent share of gross value added is used for gross labor share and investment-profit regressions. Imputed labor rent share of net value added is used for net labor share regressions.

⁺$p < 0.10$, *$p < 0.05$, **$p < 0.01$

Figure 12. Change in Gross Profitability and Imputed Labor Rent Share, by Industry, 1988–1992 to 2012–2016

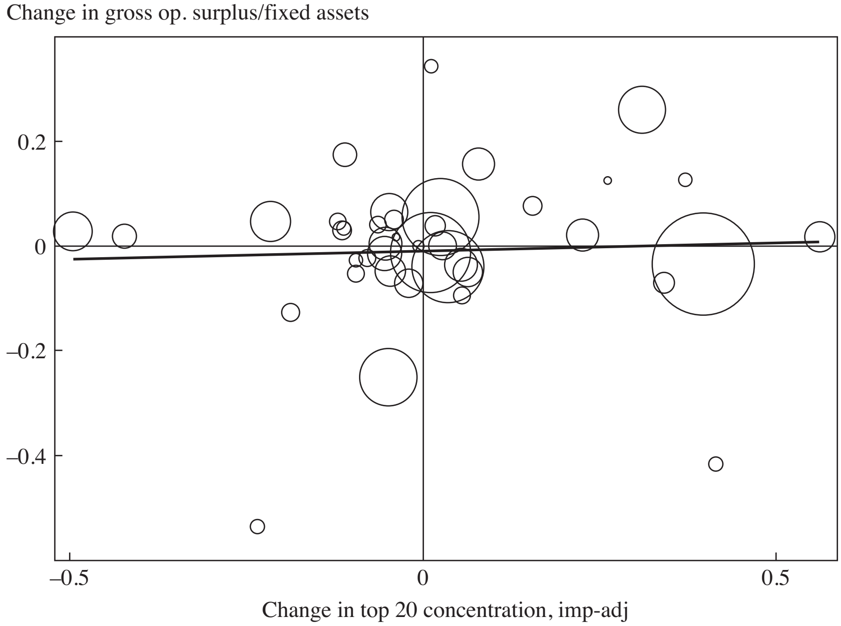


Sources: BEA industry accounts; authors' calculations.
Note: Each bubble is an industry (at BEA industry code level). Bubble size represents industry average employment over 2012–2016. The solid line is an employment-weighted line of best fit.

negative (the opposite sign than would be predicted if rising monopoly power was causing higher profitability). The coefficients from the regression over 1987–2016 with industry and year fixed effects suggest that the average increase in top twenty import-adjusted sales concentration over 1987–2016 was associated with 0.003 points increase in the median firm q at the industry level, while the average fall in the labor rent share was associated with 0.06 points increase in median firm q (although neither of these coefficients are statistically significantly different from zero in the specification with full industry and year fixed effects).[63] The median industry saw

63. The average industry's increase in import-adjusted top twenty sales concentration over 1987–2016 was 1.7 percentage points. Multiplied by the estimated coefficient of 0.16, this suggests an increase in median firm q of 0.003 points. The average industry's decline in the labor rent share over 1987–2016 was 4.6 percentage points. Multiplied by the estimated coefficient of –1.24, this suggests an increase in median firm q of 0.06 points.

Figure 13. Change in Gross Profitability and Top Twenty Sales Concentration (Import-Adjusted), by Industry, 1988–1992 to 2012–2016



Sources: BEA industry accounts; concentration data calculated from Compustat by Covarrubias, Gutiérrez, and Philippon (2019).

Note: Each bubble is an industry (at BEA industry code level). Bubble size represents industry average employment over 2012–2016. The solid line is an employment-weighted line of best fit.

an increase in its median firm q of 0.34 over the period—suggesting again that the decline in worker power has more explanatory power than the rise in concentration for changes in industry-level profitability.

IV. Unemployment and Inflation

Recent decades in the United States have seen a substantial decline in the trend unemployment rate, without inflationary pressure. The unemployment rate was below 5 percent, the level previously thought to have been the NAIRU, for nearly half of the twenty-three years from 1997 to 2020, and was below 4 percent from May 2018 until February 2020, at levels not reached since the 1960s. At the same time, inflation has been low and has shown little sign of accelerating. These facts suggest that there has been a fall in the NAIRU (Crump and others 2019; Tüzemen 2019; Blanchard, Cerutti, and Summers 2015). In this section of the paper, we argue that falling worker power could account for these broad features of the unemployment and inflation experience.

On a theoretical level, the fall in the NAIRU could be explained by a fall in worker power. Almost all models of worker insider power or rent-sharing power would predict that as worker bargaining power falls, the NAIRU would also fall. The mechanisms—and their welfare implications—vary according to the model. First, a fall in worker bargaining power may reduce the marginal cost to a firm of increasing its employment, reducing unemployment (Mortensen and Pissarides 1999; Figura and Ratner 2015). Blanchard and Giavazzi (2003) model the implications of worker power *and* monopoly power jointly; in their model falling worker power leads to lower unemployment as the incentive for firms to hire rises, while rising monopoly power leads to higher unemployment as firms reduce their output.[64] Second, this effect may be reinforced or magnified by a reduction in the distinction between insiders and outsiders in wage setting (Blanchard and Summers 1986; Calmfors and Driffill 1988; Galí 2020). Third, a reduction in the availability of high wage jobs at, for example, unionized firms may reduce the incentives for wait unemployment, where unemployed workers search for longer to try to get a high-wage job, or rest unemployment, where unemployed workers in high-rent sectors with temporary downturns wait for jobs to return (Hall 1975; Bulow and Summers 1986; Alvarez and Veracierto 1999; Alvarez and Shimer 2011).[65] Past empirical evidence suggested that areas and industries with higher rates of unionization have tended to have higher unemployment rates, and unionized firms have tended to see lower employment growth.[66] More recently, Erickson and Mitchell (2007), Figura and Ratner (2015), and Krueger (2018) have argued that the fall in labor power would lower the NAIRU, and Leduc and Wilson (2017) and Ratner

64. More specifically, their model predicts that in the short run (with no entry of firms), falling worker power reduces the labor share with no effect on unemployment, but in the long run (where all firms pay entry costs and there are no positive rents), falling worker power reduces unemployment with no effect on the labor share. If the world is always somewhere between the pure short run and pure long run—there is some entry, but there are still some positive rents—then falling worker power in their model would predict a falling labor share *and* falling unemployment.

65. On the other hand, in very frictional labor markets where a low elasticity of labor supply to the firm enables a large wage markdown, aggregate unemployment could fall as worker bargaining power rises (Manning 2003).

66. See Freeman and Medoff (1984), Summers (1986), Montgomery (1989), Blanchflower, Millward, and Oswald (1991), and Leonard (1992).

and Sim (2020) have argued that a fall in worker bargaining power could have caused the flattening of the Phillips curve.[67]

It is less clear how to reconcile trends in the NAIRU with rising globalization, technological change, or monopoly power—the other main explanations for the trends in the labor share and corporate profitability we examine in this paper. While increased globalization and technological change may have led to disinflationary pressure in the US economy, their effect on the NAIRU would be ambiguous; disinflationary pressure as a result of lower input costs may reduce the NAIRU, but the job displacement associated with both of these phenomena may increase it.[68] And it is not possible to explain the substantial fall in the NAIRU as a result of an increase in aggregate monopoly power. While theoretical models differ on whether rising monopoly power should increase unemployment or leave it constant, there is no a priori reason to believe that an increase in monopoly power would reduce unemployment, and at the same time, an increase in monopoly power may be a source of inflationary pressure.[69] Neither of these appear obviously compatible with the trends of falling unemployment and low and stable inflation that have characterized the last three to four decades (Van Reenen 2018; Basu 2019; Syverson 2019).[70]

IV.A. State-Level Evidence

The theory discussed above suggests that falling worker power could explain the aggregate decline in unemployment seen in the United States in recent decades. State-level trends in unemployment and labor rents are

67. A number of other drivers have been posited for the fall in the NAIRU, including the changing demographic composition of the workforce (Shimer 1998; Tüzemen 2019), changes in productivity growth (Ball and Mankiw 2002), improvements in job matching (Katz and Krueger 1999), and, most recently, the decline in job destruction and reallocation intensity and the aging of workers and firms (Crump and others 2019).

68. See, for example, Kohn (2005).

69. In some models of monopoly power, the employment rate is reduced with no effect on the unemployment rate. In other models, rising monopoly power leads to rising unemployment. Blanchard and Giavazzi (2003), Geroski, Gregg, and Van Reenen (1995), and Ebell and Haefke (2009), for example, show that monopoly power plus some nonzero worker bargaining power can lead to higher unemployment. Manning (1990) demonstrates that increasing returns to scale plus monopoly power can generate high unemployment equilibria. In terms of inflation, higher markups would likely imply a higher price level (in the presence of some downward nominal wage rigidity) and therefore an increase in the inflation rate during the transition from one steady state to a new, higher-markup steady state (Phelps 1968). An increase in markups, acting as a cost-push shock, would tend to imply a higher level of inflation for a given degree of labor market slack.

70. Note also that increasing monopsony power would tend to be associated with less hiring and increased labor market frictions and so also does not provide a natural explanation for a declining NAIRU.

Figure 14. State-Level Changes in Unemployment and Labor Rents, 1984–1988 to 2012–2016

Sources: CPS; authors' calculations.
Note: The solid line is a line of best fit.

consistent with this. Figure 14 shows that states with bigger falls in their imputed labor rent share over 1984–2016 also had bigger falls in their state unemployment rate.[71] Regressing the state unemployment rate on the state imputed labor rent share at the annual level, with various combinations of state and year fixed effects, we find a consistently large, positive, and significant relationship between the two variables: higher state labor rent shares are associated with higher unemployment, with the coefficient in the specification with year and state fixed effects suggesting that a 1 percentage point lower labor rent share of GDP is associated with 0.15 percentage point lower unemployment (as shown in table 5).[72]

71. The coefficient on the line of best fit is 0.36, and the p value is 0.01. The R^2 is 13 percent.

72. Disaggregating the unemployment rate by age and gender, the large, statistically significant relationship between state-level labor rents and unemployment rates holds for workers age 25–54, and 16–24, for both men and women, but not for workers age 55–65. The estimated coefficients are particularly large for all workers age 16–24 and for women age 25–54, consistent with Bertola, Blau, and Kahn's (2007) cross-national findings.

Table 5. State-Level Regressions of Unemployment on Measures of Labor Power

State-level regression of unemployment on imputed labor rent share of state GDP, 1984–2016				
Imputed labor rent share	0.14*	0.22*	0.08⁺	0.15*
	(0.06)	(0.09)	(0.04)	(0.06)
Fixed effects	None	Year	State	Year, State
Observations	1,650	1,650	1,650	1,650
State-level regression of unemployment on imputed union rent share of state GDP, 1984–2016				
Imputed union rent share	0.56**	0.60*	0.50**	0.54*
	(0.18)	(0.24)	(0.16)	(0.24)
Fixed effects	None	Year	State	Year, State
Observations	1,650	1,650	1,650	1,650

Source: Authors' calculations.
Note: Robust standard errors, clustered at state level, in parentheses.
⁺$p < 0.10$, *$p < 0.05$, **$p < 0.01$

IV.B. Industry-Level Evidence

Industry-level patterns in unemployment and labor rents are also consistent with the hypothesis that declining worker power has lowered the NAIRU. As we found at the state level, industries that saw larger declines in their imputed labor rent share saw larger declines in their industry-level unemployment rate (figure 15).[73] Regressions of the annual industry-level unemployment rate on the imputed labor rent share and imputed union rent share, with industry and year fixed effects, have positive and significant coefficients (table 6), with the magnitude in the specification with industry and year fixed effects suggesting that a 1 percentage point lower imputed labor rent share is associated with a 0.1 percentage point decline in industry unemployment.[74]

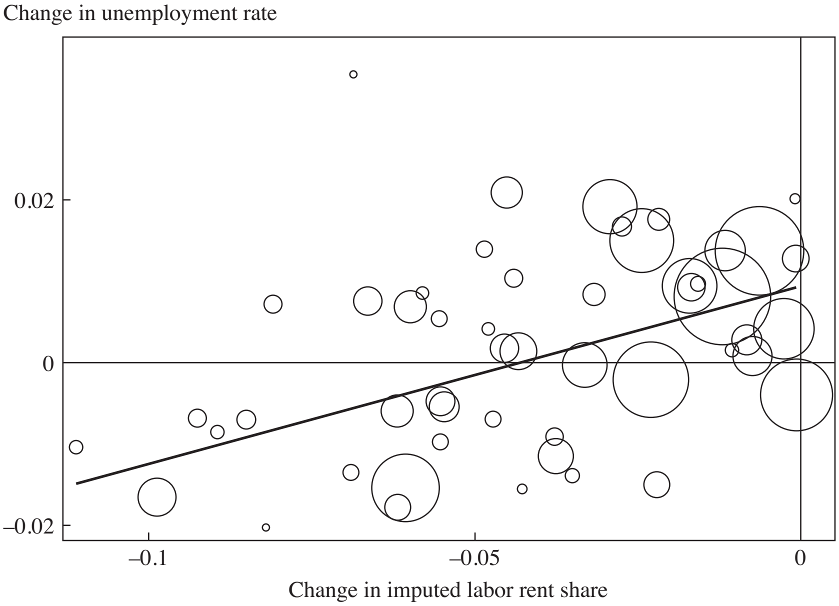
IV.C. Unemployment for College-Educated and Non-College-Educated Workers

In section II.D, we decomposed the decline in labor rents for workers with and without a college degree (bachelor's or more) over 1984–2016 and showed that while both groups saw a decline in their labor rents,

73. We measure industry unemployment in the CPS, defining it as the unemployment rate among all workers who reported having worked in a given industry in their current job (if employed) or most recent job (if unemployed).

74. Supplementing this analysis, we also show in online appendix C.9 that there is a significant relationship between industry-level unemployment and unionization rates and between industry-level labor market tightness, labor rent shares, and unionization rates. Note that, in contrast, regressions of the annual industry-level unemployment rate on measures of industrial concentration show no significant relationship, and the coefficients are positive.

Figure 15. Change in Unemployment and Imputed Labor Rent Share, by Industry, 1988–1992 to 2012–2016



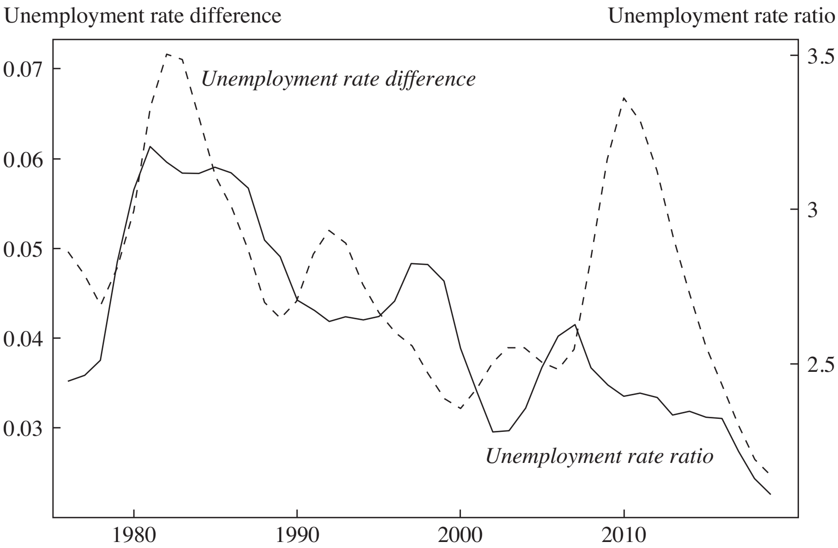
Sources: CPS; authors' calculations.
Note: Each bubble is an industry (at the BEA industry code level), where the size of the bubble represents industry average employment over 2012–2016. The solid line is an employment-weighted line of best fit.

Table 6. Industry-Level Regressions of Unemployment on Measures of Labor Power

Industry-level regression of unemployment on imputed labor rent share of gross value added, 1987–2016 (fifty-one industries)				
Imputed labor rent share	–0.16**	–0.16**	–0.03	0.10**
	(0.05)	(0.05)	(0.03)	(0.03)
Fixed effects	None	Year	Ind	Year, Ind
Observations	1,530	1,530	1,530	1,530
Industry-level regression of unemployment on imputed union rent share of gross value added, 1984–2016 (fifty-one industries)				
Imputed union rent share	–0.27*	–0.23+	–0.21*	0.20**
	(0.10)	(0.12)	(0.08)	(0.06)
Fixed effects	None	Year	Ind	Year, Ind
Observations	1,530	1,530	1,530	1,530

Source: Authors' calculations.
Note: Robust standard errors, clustered at industry level, in parentheses.
+$p < 0.10$, *$p < 0.05$, **$p < 0.01$

Figure 16. Relative Unemployment Rates of Workers with and without a College Degree, 1976–2019



Source: CPS.

Note: Unemployment rate difference is defined as unemployment rate for workers without a bachelor's degree ("non-college") minus unemployment rate of workers with a bachelor's degree or better. The ratio is defined as non-college-educated unemployment rate divided by unemployment rate of workers with a bachelor's degree or better. Data points are three-year moving averages.

the decline was substantially larger for non-college-educated workers. If declining labor rents leads to a lower NAIRU, one might expect to see larger declines in average unemployment for non-college-educated workers than for college-educated workers over the same period. This has been the case: the unemployment rate of workers without a four-year college degree has fallen substantially relative to the unemployment rate of workers with a bachelor's degree, as shown in figure 16.

IV.D. Quantitative Implications for the NAIRU

Can we say anything about whether the magnitude of the decline in worker power is big enough to account for the decline in the NAIRU? One recent study on this topic is Figura and Ratner (2015), who study the decline in worker power as proxied for by the decline in the labor share of income. They show that industries and states with bigger falls in their labor share over 2001–2014 saw bigger increases in their vacancy-to-unemployment ratio (labor market tightness). They argue that this is consistent with a decline in worker bargaining power increasing the incentive for firms to create jobs and that the decline in the labor share of income could have led to a two-thirds of a percentage point fall in the NAIRU.[75] We can similarly use our state-level and industry-level estimates to back out a naive extrapolation of the aggregate relationship between worker power and unemployment. Applying the coefficients from the state-level regressions in table 5 to our estimate of the fall in labor rents in the nonfinancial corporate sector over 1982–2016 (a fall of 5.1 percentage points) would have predicted a fall of three quarters of a percentage point in the NAIRU.[76] We have reason to believe that both the Figura and Ratner (2015) estimate and our estimate of the effect of the decline of worker power on the NAIRU may be underestimates of the true effect, since they are based on state- and industry-level variation which may miss some aggregate effect, and since the imperfection of the labor share (in the case of Figura and Ratner) or the imputed labor rent share (in our case) as proxies for the decline in worker power is likely to cause attenuation bias.[77]

V. Possible Objections and Further Considerations

In this section, we examine trends in aggregate investment, firm-level labor shares and markups, rising top-end labor compensation, and occupational licensing in light of the declining worker power hypothesis. We also evaluate the evidence for alternative explanations of the declining labor share and rising corporate valuations—globalization, technological

75. More formally, they argue that the negative relationship they find between the labor share and the vacancy-to-unemployment ratio is consistent with a counterclockwise rotation in the job creation curve in a standard DMP search model. After estimating the slope of the Beveridge curve, they can then estimate the degree to which a decline in worker bargaining power may affect equilibrium unemployment.

76. Our state-level estimates of the labor rent share—which we use to generate the estimated relationship between changes in the labor rent share and unemployment—calculate labor rents as a share of private sector value added. In this calculation we use our estimate of the decline in labor rents as a share of nonfinancial corporate value added. This calculation therefore implicitly assumes that the decline in the labor rent share of gross value added in the financial sector and in the nonfinancial noncorporate sector was also 5.1 percentage points. Our estimates of labor rents in the entire corporate sector (including finance) were very similar to our estimates of labor rents in the nonfinancial corporate sector; they can be found in online appendix B.2.

77. While a full model-based investigation of the degree to which the decline in worker power may have affected the NAIRU is beyond the scope of this paper, we carry out four back-of-the-envelope exercises in online appendix E. These illustrate that, in simple models with plausible parameter values, it is possible for the decline in worker power that we have seen to generate very large changes in the NAIRU.

change, and rising monopoly power—in light of cross-national and cross-industry evidence.

V.A. Investment

Investment has been falling over recent decades relative to measures of corporate profitability such as operating surplus and Tobin's q, as well as relative to GDP and fixed assets (Gutiérrez and Philippon 2017; Alexander and Eberly 2018; Crouzet and Eberly 2019). These trends have been a major motivator of the monopoly power argument (Gutiérrez and Philippon 2017; Eggertsson, Robbins, and Wold 2018). One might argue these trends in investment are hard to reconcile with our argument that there has been a macroeconomically important decline in worker power: some models predict that a decline in worker power, reducing the marginal cost of production, would lead to an increase in investment.[78] To what extent are the facts on investment compatible with our argument of declining worker power?

First, we note that it is not clear that investment, properly measured, *has* declined substantially relative to value added or fixed assets. The relative price of investment goods has declined, meaning that while there has been a decline in net investment relative to net value added in nominal terms, there has been no decline in net real fixed investment relative to net real value added in the nonfinancial corporate sector (as shown in figure 17).[79] And Crouzet and Eberly (2019, 2020) show that a rise in intangible investment could account for the majority of the apparent decline in investment relative to fixed assets.

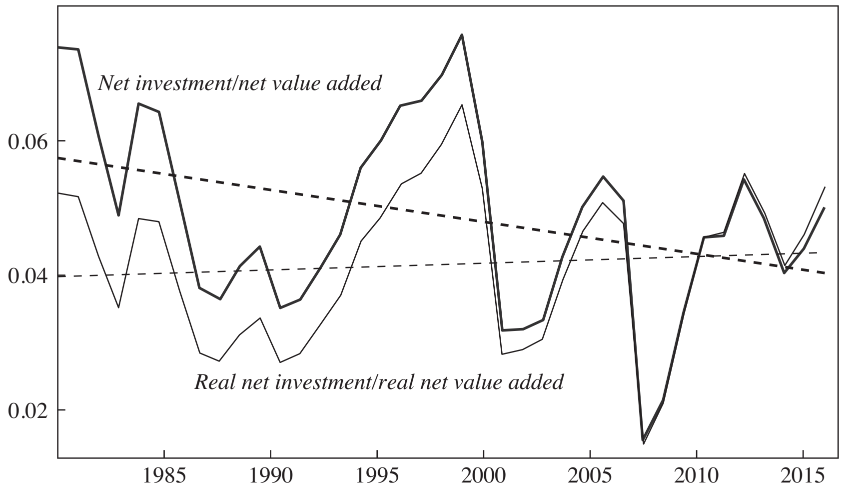
Second, we note that the theoretical predictions of declining worker power for investment are actually ambiguous. It is possible that a decline in worker power leads to less investment: by reducing the marginal cost of labor to firms, declining worker power may lead to the substitution of labor for capital (or at least, less substitution of capital for labor), reducing investment relative to a scenario where worker power had not declined.

Third, the fall in investment relative to measures of corporate profits can be explained by our declining worker power hypothesis. In efficient bargain

78. As argued by Eggertsson, Robbins, and Wold (2018), for example.

79. Net investment to net value added is calculated using data on gross nonresidential investment and the consumption of nonresidential fixed capital by nonfinancial corporate business, from the Federal Reserve's Financial Accounts of the United States, Z.1, and gross value added in the nonfinancial corporate business sector from BEA NIPA. For the ratio of real net investment to real net value added, investment is deflated by the implicit price deflator for nonresidential fixed private sector domestic investment from the BEA, and value added is deflated by the implicit price deflator for nonfinancial corporate business from the Bureau of Labor Statistics.

Figure 17. Real and Nominal Net Investment over Net Value Added, Nonfinancial Corporate Sector



Sources: Federal Reserve Financial Accounts of the United States, Z.1; BEA NIPA.

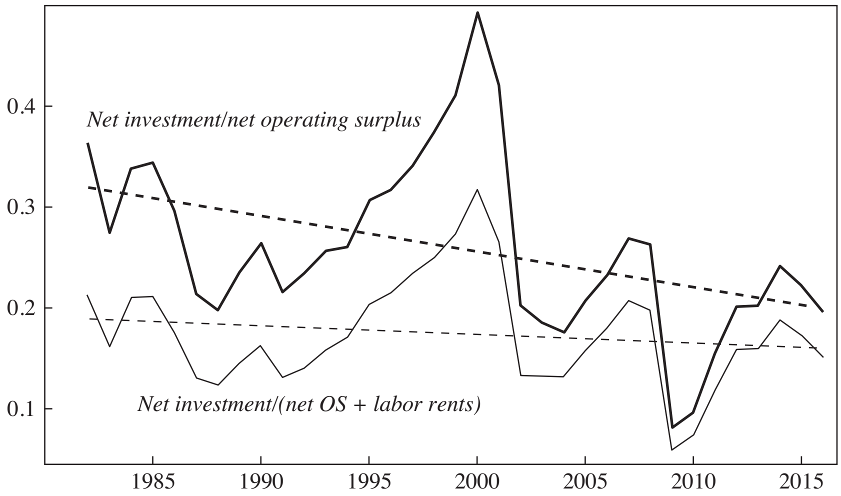
models of worker rent-sharing (our model in section III.A, for example) the degree of worker power does not affect the firm's investment decision. The firm makes its investment decisions in order to maximize total profits, then distributes the rents between labor and capital. To understand if investment has fallen relative to the underlying profitability of firms, we must therefore measure both profits to capital and profits to labor. Defining the ratio of investment to total profits as follows:

$$\frac{\text{Investment}}{\text{Total profits}} = \frac{\text{Investment}}{\text{Net operating surplus} + \text{imputed labor rents}}$$

we show in figure 18 that while net investment over net operating surplus (profits *to capital*) has fallen substantially over the last thirty years in the nonfinancial corporate sector, average net investment over our measure of net *total profits* has declined only very slightly. That is, even nominal investment has not weakened much relative to our measure of firms' total profitability.[80] The relationship between labor power and investment-to-profits

80. Crouzet and Eberly (2020) attribute a share of the growing weakness of investment relative to Tobin's q to product market rents. Our explanation could be compatible with this: instead of the product market rents arising from increased monopoly power, they may have been rents that were previously paid to labor so did not show up in Tobin's q.

Figure 18. Net Investment to Profits to Capital and Imputed Total Profits, Nonfinancial Corporate



Sources: Federal Reserve Financial Accounts of the United States, Z.1, Flow of Funds Accounts; BEA NIPA; authors' calculations.

Note: Investment is measured as gross fixed investment in nonresidential structures, equipment, and intellectual property products for nonfinancial corporate business. We obtain net investment by subtracting the consumption of fixed capital for the nonfinancial corporate sector from gross investment. The labor rents measure is constructed as described in section II.

also holds at the industry level: industries with larger declines in their imputed labor rent share saw larger declines in the ratio of investment to operating surplus, even in annual regressions when controlling for a variety of industry and year fixed effects (table 4).[81]

V.B. Firm-Level Dynamics: Labor Shares and Markups

Our analysis in this paper is primarily at the industry and aggregate level. Recent research has emphasized the role of firm-level dynamics in trends in labor shares, markups, and wages. First, several papers find a large role for between-firm reallocation in the decline of the labor share and rise in measured markups. Second, research with matched employer-employee data suggests that the dispersion of average earnings at the firm

81. In contrast, coefficients on average top twenty sales concentration are noisy (see table 4), and there is no apparent relationship between the change over 1988–2016 in the average top twenty sales concentration ratio and the ratio of investment to gross operating surplus (see online appendix C.10).

level has risen. Can we reconcile our results with this firm-level evidence on labor shares, markups, and wages?

LABOR SHARE AND MARKUPS Autor and others (2020) find that two-thirds of the decline in the aggregate labor share can be explained by between-firm reallocation, with one-third explained by within-firm falls in the labor share. The median firm saw no decline in their labor share, while firms with initially low labor shares saw their labor shares fall still further. Kehrig and Vincent (2020) find similar dynamics in manufacturing, showing that the decline in the labor share is driven by establishments which are growing in size and at the same time see falling labor shares. De Loecker, Eeckhout, and Unger (2020) find that the rise in the aggregate measured markup results largely from a reallocation of activity to high-markup firms, the median markup did not change, and markups for already high-markup firms increased.

It is clear that our proposed mechanism—a fall in labor rent-sharing power—could explain *within-firm* declines in labor shares and increases in measured markups. It is also possible to reconcile our proposed mechanism with the portion of the decline in the labor share (or rise in measured markups) that results from the *reallocation* of economic activity across firms. First, it could simply be the case that firms which experienced bigger falls in worker power also grew faster for some exogenous reason. Second, it is possible that this faster growth itself is at least partly a result of falling worker power. To see this, note that if workers receive a competitive wage plus some portion of a firm's rents, then unit labor costs are higher at high-rent firms than at low-rent firms, but unless workers' share of rents in high-rent firms is higher than the aggregate labor share, high-rent firms will still have lower labor shares than low-rent firms. Therefore, as workers' rent-sharing power declines, unit labor costs fall disproportionately more at high-rent, low-labor-share firms than at low-rent, high-labor-share firms. This improves the competitive advantage of high-rent firms, creating an incentive for them to expand. This would lead to a reallocation of economic activity from high-labor-share to low-labor-share firms.

FIRM WAGE EFFECTS There has been an increase in the dispersion of average wages at the firm level over recent decades, which has led to suggestions that this could indicate a divergence in firm-level rents (Barth and others 2016). This might be seen as supporting the hypothesis of rising monopoly power, rather than declining worker power. In fact, the evidence is more consistent with declining worker power. Song and others (2019) use matched employer-employee data to decompose the variance of US wages into firm effects, worker effects, and the covariance of the two, following Abowd, Kramarz, and Margolis (1999). The firm effects indicate the firm-specific pay premium, holding worker quality constant, and can be interpreted as some combination of rent-sharing and compensating differentials (Sorkin 2018; Card and others 2018). Song and others (2019) show that the increase in the variance of firm-level average wages over 1980–2013 was entirely due to an increase in the *sorting* of high-wage workers into high-wage firms and not an increase in the dispersion of the firm premia paid to equivalent workers. In fact, they find a small decline in the variance of firm effects over the period. These trends are consistent with a decline in rent-sharing: the decline in the variance of firm fixed effects could reflect declining wage premia in formerly high-wage firms, and the increase in the sorting of high-wage workers into high-wage firms (and vice versa) could reflect the fissuring of the workplace. On the other hand, if an increase in monopoly power had caused total rents to increase, holding constant the initial degree of rent-sharing with workers, one would have expected firm effects to become more dispersed rather than less (if rents increased more for already high-rent firms).

Note that the decline in the variance of *firm* fixed effects estimated by Song and others (2019) has been substantially smaller than the decline we estimate in the variance of *industry* fixed effects. There are two ways to reconcile this. First, note that a large decline in the variance of industry wage premia, but a small decline in the variance of firm wage premia, would be consistent with an aggregate decline in labor rents as a result of the fissuring of the workplace, as an increasing share of workers work at firms with low rents (and fewer at firms with high rents).[82] Second, evidence from Lachowska and others (2020), who carry out an AKM decomposition in Washington State over 2002–2014, suggests that the underlying secular decline in the variance of firm fixed effects over recent decades may have been larger than that estimated by Song and others (2019), both because the decline in the variance of firm fixed effects for *hourly wages* may have been larger than that for annual earnings (where Song and others focus on the latter), and because the endpoint of the analysis by Song and others (2019) appears to have been at a point

82. Following the suggestion of Christina Patterson in her remarks at the Spring 2020 BPEA meetings, we note that the relationship between firm and industry effects can be written as $\gamma_{ind} = \sum_j \frac{E_j}{E_{ind}} \gamma_j$, where γ_{ind} and γ_j denote industry and firm wage effects, respectively, and E_{ind} and E_j denote industry and firm employment, respectively.

where the variance of firm fixed effects may have been particularly high for cyclical reasons.[83]

V.C. Labor Rents to the Highly Paid: Executive Compensation and Finance

There has been roughly a doubling of the share of national income accruing to executives, managers, and supervisors in nonfinancial firms since 1979 (Bakija, Cole, and Heim 2012). High-earning financial sector workers have also seen large rises in their compensation. Could these reflect rising labor rents?[84]

First, note that we estimate labor rents from the CPS, where the earnings data are top-coded and nonresponse is high for people in the top tail of the income distribution. This means that our estimate of the decline in labor rents should be considered to be the decline in rents for the majority of workers but not including the highest paid—and so, not including top executives, managers, or many financial sector workers.[85]

It is therefore plausible that some of the lost labor rents we measure were redistributed to top management and executives, rather than to shareholders. Indeed, this could be consistent with our evidence, since we estimate that the decline in the labor rent share of value added in the nonfinancial corporate sector (for the majority of workers) was greater than the actual decline in the labor share (which includes executive compensation). Note, though, that the majority of the increase in executive compensation (as a share of total income) over this period accrued to executives and managers who receive self-employment, S-corporation, or partnership income (Bakija,

83. Song and others (2019) find that over 1980–87 to 2007–13 there was a decline of about 3.5 percent in the variance of firm fixed effects. While they use different data sets, Lachowska and others (2020) and Song and others (2019) find very similar declines in the variance of firm fixed effects for annual earnings over the period they study in common (2002–2014), suggesting the two studies may be comparable. Lachowska and others (2020) find a much larger decline in the variance of firm fixed effects for hourly wages than for annual earnings over this period, and they find large countercyclicality in the variance of firm fixed effects; their estimates suggest that the variance of firm fixed effects will have been particularly high during the 2007–2013 period, the endpoint of the comparison in Song and others (2019).

84. See Bivens and Mishel (2013) for evidence on the existence of rents in executive pay, and Philippon and Reshef (2012) for evidence on rents in financial sector compensation.

85. Specifically, our baseline estimate of labor rents for the nonfinancial corporate sector will omit high-paid executives and managers because of CPS top-coding or nonresponse and will omit all financial sector workers by construction. Our estimate of labor rents for the entire corporate sector will include many financial sector workers but will omit increases in pay for the highest-paid financial sector workers because of CPS top-coding. See online appendix B.2 for more details.

Cole, and Heim 2012; Smith and others 2019).[86] Since it is ambiguous whether income from these sources should be considered capital or labor income, it is unclear whether to consider the rising income of executives and managers of S-corporations and partnerships as a redistribution of rents from workers' labor income to managers' labor income or simply from labor to capital.[87]

It is also plausible that some of the lost labor rents we measure were redistributed to high-paid financial sector workers. When estimating labor rents for the entire corporate sector (including finance), we find a very similar decline in labor rents as we do for the nonfinancial sector—meaning that the inclusion of the majority of financial sector workers does not affect our conclusions. However, since the CPS earnings data are top-coded, our calculation will miss any increase in rents accruing to very highly paid professionals in finance. In our CPS-ORG data, the share of workers in finance, insurance, and real estate who had top-coded earnings rose from 2 percent in 2000 to 9 percent by 2019. It is possible that these workers saw their labor rents increase over the period where the majority of workers saw labor rents decrease—but note that since this is a relatively small group of workers, even rather drastic increases in rents for the top 5–10 percent of financial sector workers would not have made a major difference in the overall trend in labor rents for the entire corporate sector.[88]

V.D. Occupational Licensing

While unionization, industry wage premia, and firm size wage premia have fallen over recent decades, the extent of occupational licensing has

86. Bakija, Cole, and Heim (2012) estimate that the increase in the income share of managers, executives, and supervisors in the top 1 percent of all income earners who work for closely held businesses was around 2.2 percentage points over 1979–2005, while the increase in the income share of salaried managers, executives, and supervisors in the top 1 percent of all income earners was only around 0.4 percentage points.

87. Note also that Smith and others (2019) argue that the decline in the labor share has been overstated because of the increase in income accruing to the top 1 percent of earners which comes from pass-through enterprises. This income is booked as capital income but some may more appropriately be considered labor income. While the degree of the decline in the aggregate labor income share may be ambiguous as a result of the difficulties of imputing pass-through income to labor or capital (and imputing self-employment income), what is *not* ambiguous is that the share of total income going to the vast majority of workers has declined since the 1980s. For example, Piketty, Saez, and Zucman (2018) estimate that for the bottom 99 percent of people, for example, the share of total national income accounted for by labor compensation declined from 69 percent in 1978 to 59 percent in 2014.

88. See online appendix B.2 for our estimates of labor rents in the entire corporate sector and a discussion of the effect of top-coding of earnings on our calculation of labor rents in finance.

risen. Have we overestimated the decline in labor rents by failing to consider occupational licensing?

We believe that accounting for the rise of occupational licensing would not substantially change our results. First, note that for many professions in which occupational licensing has increased in recent years, occupational licenses are less likely to transfer rents from workers to capital owners than they are to transfer rents from unlicensed workers to licensed workers or from consumers to workers (for example, hairdressers, manicurists, and cosmetologists, real estate agents, or self-employed workers in the building trades). Recent work by Kleiner and Soltas (2019) estimates that 70 percent of the welfare loss of marginal occupational licensing is borne by workers. Even if we were to assume that *all* rents accruing to workers as a result of occupational licensing were obtained at the expense of capital, a back-of-the-envelope calculation suggests that the rise of occupational licensing could only have resulted in an increase in labor rents of 0.2–0.7 percentage points of value added; the share of the US labor force required to have an occupational license is estimated to have risen by around 7–12 percent from the 1980s to 2008 (Kleiner and Krueger 2013; CEA 2016), and the wage premium for licensed workers in the United States appears to be in the range of 4–8 percent (Gittleman, Klee, and Kleiner 2018; Bryson and Kleiner 2019).

V.E. Further Evidence on Alternative Hypotheses for Trends in Labor Shares and Corporate Profits

In this section, we address a number of empirical trends which point to weaknesses in the arguments that globalization, technological change, or monopoly and monopsony power were the predominant drivers of the falling labor share and rising corporate profits.

First, we note that while technological change and globalization are ubiquitous, the extent of increases in inequality—both between capital and labor incomes and within labor incomes—differ substantially across countries (Gutiérrez and Piton forthcoming). This would tend to suggest a substantial role for country-specific factors in explaining the decline in the labor share—as argued by Philippon (2020) among others—pointing up the monopoly power or worker power explanations as candidates.

A large proportion of the decline in the US labor share can be accounted for by the manufacturing sector. The centrality of the manufacturing sector in the decline in the US labor share would tend to favor the declining worker power hypothesis over the rising monopoly power hypothesis; given the increases in international trade driven by the opening of low-wage economies to international markets and reductions in transport costs and trade barriers, it seems unlikely that US manufacturing has seen a substantial increase in product market power over recent decades. In contrast, the manufacturing sector saw large declines in unionization over recent decades and can account for a large share of our estimated decline in labor rents.

Our hypothesis, which emphasizes the relative power of labor and capital, can therefore fit the combination of cross-country and cross-industry facts better than hypotheses based on globalization, technological change, or monopoly power (given far more empowered shareholders and weaker unions in the United States than in the rest of the industrial world). In keeping with this, cross-country evidence from Kristal (2010) and Jaumotte and Osorio Buitron (2015) suggests that countries with bigger declines in unionization saw bigger declines in their labor shares and bigger increases in income inequality.[89]

Second, while monopoly power and monopsony power are without doubt present in certain parts of the US economy—and our baseline framework in fact assumes the existence of both types of power—we also note that the direct evidence of a large aggregate increase in either monopoly power or monopsony power is unclear.[90]

The large rise in industry-level sales concentration over recent decades has frequently been invoked as a likely driver of rising monopoly power (Grullon, Larkin, and Michaely 2019; Gutiérrez and Philippon 2017, 2019). Yet industrial organization economists point up a number of reasons to be skeptical that this increase in concentration reflects a large increase in aggregate monopoly power (see, for example, Shapiro 2018; Berry, Gaynor, and Scott Morton 2019; Basu 2019; Syverson 2019). First, it is

89. Bental and Demougin (2010) also argue that cross-country trends in the labor share may have been driven by an erosion of worker bargaining power, but as a result of improved monitoring technologies. Earlier work studying cross-country trends in labor shares includes Bentolila and Saint-Paul (2003).

90. Several recent studies demonstrate the presence of monopoly and monopsony power in the US economy. In terms of monopoly power, Covarrubias, Gutiérrez, and Philippon (2019) and Philippon (2020), for example, document that since 2000, rising concentration has been associated with slower turnover of lead firms and rising prices, particularly in telecoms, airlines, and banking; they present case studies of several products where prices are substantially higher in the United States than in Europe. In terms of monopsony power, Berger, Herkenhoff, and Mongey (2019) estimate welfare losses of 5 percent of lifetime income arising from employers' power in the labor market (as indexed by workers' elasticity of labor supply); and Schubert, Stansbury, and Taska (2020) and Arnold (2020) find sizeable negative effects on wages for workers in highly concentrated labor markets. See Sokolova and Sorensen (2020) for a review of the empirical evidence on the elasticity of labor supply to the firm.

not clear whether this large aggregate increase is still present when defining markets appropriately; import-adjusted measures of sales concentration in manufacturing have fallen or risen only marginally since the 1980s (Covarrubias, Gutiérrez, and Philippon 2019), and in many service industries, where the relevant market is often smaller than the entire US market, local-level sales concentration is actually falling, possibly even reflecting increased local-level competition as large firms spread their business into new markets (Rossi-Hansberg, Sarte, and Trachter forthcoming). Second, industries may become more concentrated as efficient firms compete and win market share: several authors have documented a relationship between rising product market concentration and rising productivity in certain sectors (Peltzman 2018; Autor and others 2020; Ganapati forthcoming; Crouzet and Eberly 2019). Third, even where concentration ratios have increased in well-defined markets, they are usually below levels which typically raise profit concerns (Shapiro 2018).[91] In addition, we note that the substantial decline in the large-firm wage premium is the opposite of what one would expect to see if large firms were gaining more monopoly power.

Similarly, there is less direct evidence of a rise in labor market monopsony power—in terms of an increasingly inelastic labor supply curve to firms—than there is of a fall in worker power. It does not seem plausible that monopsony power has increased as a result of an increase in labor market concentration (Bivens, Mishel, and Schmitt 2018); local labor market concentration is low for most workers, particularly when considering the availability of jobs in other occupations or industries (Schubert, Stansbury, and Taska 2020) and has actually fallen, not risen, for most workers over recent decades (Rinz 2018). Berger, Herkenhoff, and Mongey (2019) estimate that the fall in local labor market concentration since the 1970s was large enough to predict a 3 percentage point *increase* in the labor share. And while the proliferation of noncompete clauses and occupational licensing requirements may have increased switching costs for some workers, the rise of the Internet at the same time should have substantially reduced the costs of job searches for workers and employers,

91. In recent years some authors have also argued that the rise in common ownership across firms, as documented by Azar, Schmalz, and Tecu (2018) among others, has led to reduced competition and increased monopoly power (Azar and Vives 2019). More research would be valuable in this regard; the theoretical links between common ownership concentration and monopolistic behavior by firms remain debated, and there does not yet appear to be a clear empirical consensus on the relationship between common ownership and industry-level outcomes like investment, prices, markups, and production (Schmalz 2018; Backus, Conlon, and Sinkinson 2019).

so the net change in the degree of labor market frictions is unclear.[92] One piece of evidence which might indicate a rise in monopsony power is Webber (2018), who estimates a decline in the firm-level elasticity of quits to the wage over 2003–2011; more research would be valuable to identify whether this reflects a long-term trend or the slow labor market recovery following the Great Recession.

VI. Concluding Remarks

The evidence in this paper suggests that the American economy has become more ruthless, as declining unionization, increasingly demanding and empowered shareholders, decreasing real minimum wages, reduced worker protections, and the increases in outsourcing domestically and abroad have disempowered workers—with profound consequences for the labor market and the broader economy. We argue that the reduction in workers' ability to lay claim to rents within firms could explain the entirety of the change in the distribution of income between labor and capital in the United States in recent decades and could also explain the rise in corporate valuations, profitability, and measured markups, as well as some of the decline in the NAIRU. We believe the declining worker power hypothesis has been substantially underemphasized as a cause of these macroeconomic trends, relative to other proposed causes: globalization, technological change, and rising monopoly or monopsony power.

An important set of issues which we do not explore in detail relate to inequality in labor income. It seems plausible that the same kinds of situations that encourage rent-sharing also encourage the compression of compensation relative to productivity: unions, generous benefit structures, formalized wage-setting processes, and so forth. Consistent with this, in section II.D we find that the decline in labor rents has been greater for workers without college degrees than for those with college degrees.[93] It is also plausible that the decline in the rent-sharing power of the majority

92. On noncompete clauses and no-poaching agreements, see Kleiner and Krueger (2013), Krueger and Ashenfelter (2018), Furman and Krueger (2016), and Starr, Prescott, and Bishara (2019). On the Internet and job searches, see Stevenson (2009), Kuhn and Mansour (2014), and Bhuller, Kostøl, and Vigtel (2020). There has been a decline in the job-switching rate over time; this may either suggest an increase in the costs of job switching, consistent with higher monopsony power, or a decrease in the dispersion of job-specific rents, reducing workers' incentive to switch jobs (Molloy and others 2016).

93. There is a large body of work consistent with this. Several authors document an important role for declining unionization in the rise in wage inequality (DiNardo, Fortin, and Lemieux 1996; Card 1996; Farber and others 2018); others document a role for the rise in outsourcing and the fissuring of the workplace (Weil 2014).

of workers could explain some of the increase in the income share of the top 1 percent. Over 1979–2014, the income share of the top 1 percent is estimated to have risen by between 4.9 and 9 percentage points (Auten and Splinter 2019; Piketty, Saez, and Zucman 2018). If we assume that all of the decline in labor rents we estimate in this paper represented redistribution from the bottom 99 percent to the top 1 percent (whether as labor or capital income), it could explain between 41 percent and 76 percent of the entire increase in the top 1 percent income share over the last forty years. If we assume instead that labor rents were redistributed as capital income across the entire income distribution, but in proportion to the actual distribution of capital income arising from firm ownership, then our estimated decline in labor rents could still account for 24–45 percent of the increase in the income share of the top 1 percent.[94]

In future research it would be valuable to more explicitly consider alternative bargaining models and their implications for wages and employment and for total output and investment. A further promising avenue is distinguishing between the degree of product market monopoly power versus labor market power in the US economy by estimating markups on different types of inputs. With sufficiently detailed data on input costs, markups could be estimated on nonlabor inputs and on labor inputs separately. Markups over labor and nonlabor inputs following the same path would be consistent with a rise in monopoly power; markups over nonlabor inputs staying constant while markups over labor rise would be more consistent with a fall in worker power or a rise in monopsony power.[95]

A fair question about the labor rents hypothesis regards what it says about the secular stagnation hypothesis that one of us has put forward (Summers 2013). We believe that the shift toward more capital income, which occurs as labor rents decline, operates to raise saving and reduce demand. The impact on investment of reduced labor power seems to us ambiguous, with lower labor costs encouraging expanded output on the one hand and on the other encouraging more labor-intensive production, as discussed in section V. So decreases in labor power may operate to promote the reductions in demand and a rising gap between private saving and investment that are defining features of secular stagnation.

Finally, it is worth highlighting that the declining worker power hypothesis is perhaps more deeply threatening to existing thinking than the other prominent hypotheses for the causes of the decline in the labor share. The

94. For details of our calculations, see online appendix C.11.

95. However, finding differential trends in markups on labor inputs versus nonlabor inputs would not be conclusive evidence because this could also be driven by technological change (Baqaee and Farhi 2020).

globalization or technological change perspectives would imply that any adverse distributional consequences have come alongside greater efficiency, which would have made Pareto-improving redistribution possible (at least in principle). The monopsony and monopoly perspectives suggest that the rise in inequality has come alongside the economy becoming less efficient, which puts economists in the congenial position of arguing for policies that simultaneously perfect markets, increase efficiency, and promote fairness. In contrast, the declining worker power perspective would imply that the increased inequality we have seen over recent decades may not have come alongside greater efficiency. And the policy implication if these trends are to be reversed—doing more to preserve rent-sharing—interferes with pure markets and may not enhance efficiency on at least some measures.[96]

More profoundly, if the decline in worker power has been a major cause of increases in inequality and lack of progress in labor incomes, if policy-makers wish to reverse these trends, and if these problems cannot be addressed by making markets more competitive, this raises questions about capitalist institutions. In particular, it raises issues about the effects of corporate governance arrangements that promote the interests of shareholders only versus a broader set of stakeholders—a constantly simmering debate that has gained new prominence with the Business Roundtable's embrace of stakeholder capitalism. And it suggests that institutions that share rents with workers are likely to be necessary as a form of countervailing power, of the sort initially proposed by Galbraith (1952).

ACKNOWLEDGMENTS We thank the editors, Jan Eberly and Jim Stock, and discussants Steven Davis and Christina Patterson for their comments. For helpful discussions and comments we thank Paweł Bukowski, Gabriel Chodorow-Reich, Ray Fair, Richard Freeman, John Haltiwanger, Kathryn Holston, Larry Katz, Morris Kleiner, Pat Kline, Thomas Kochan, Larry Mishel, Simon Mongey, Peter Norlander, Valerie Ramey, Jake Rosenfeld, Bob Solow, and the participants of the Spring 2020 Brookings Papers on Economic Activity conference, the Spring 2020 Harvard macroeconomics workshop, and the Spring 2020 Harvard econometrics clinic. We are very grateful to Germán Gutiérrez and Thomas Philippon, David Baqaee and Emmanuel Farhi, Nicholas Bloom and David Price, Matthias Kehrig, Christina Patterson, and Maria Voronina for providing us with data.

96. The degree to which labor market rent-sharing institutions promote or reduce aggregate efficiency depends on the underlying degree of competition in the labor market, the availability of rents in the product market, and the nature of the rent-sharing institutions, as discussed by Manning (2003) and others.

References

Abdih, Yasser, and Stephan Danninger. 2017. "What Explains the Decline of the US Labor Share of Income? An Analysis of State and Industry Level Data." Working Paper WP/17/167. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2017/07/24/What-Explains-the-Decline-of-the-U-S-45086.

Abowd, John M. 1989. "The Effect of Wage Bargains on the Stock Market Value of the Firm." *American Economic Review* 79, no. 4: 774–900.

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67, no. 2: 251–333.

Abowd, John M., Francis Kramarz, Paul Lengermann, Kevin L. McKinney, and Sebastien Roux. 2012. "Persistent Inter-industry Wage Differences: Rent Sharing and Opportunity Costs." *IZA Journal of Labor Economics* 1, no. 7: 1–25.

Abowd, John M., Paul Lengermann, and Kevin L. McKinney. 2003. "The Measurement of Human Capital in the US Economy." Working paper. Suitland, Md.: US Census Bureau.

Abowd, John, Kevin L. McKinney, and Nellie Zhao. 2018. "Earnings Inequality and Mobility Trends in the United States: Nationally Representative Estimates from Longitudinally Linked Employer-Employee Data." *Journal of Labor Economics* 36, no. S1: S183–S300.

Acemoglu, Daron, Phillipe Aghion, and Giovanni L. Violante. 2001. "Deunionization, Technical Change, and Inequality." *Carnegie-Rochester Conference Series on Public Policy* 55, no. 1: 229–64.

Acemoglu, Daron, and Pascual Restrepo. 2018. "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review* 108, no. 6: 1488–542.

Ahlquist, J. S. 2017. "Labor Unions, Political Representation, and Economic Inequality." *Annual Review of Political Science* 20:409–32.

Akerlof, George A., and Janet L. Yellen, eds. 1986. *Efficiency Wage Models of the Labor Market*. New York: Cambridge University Press.

Alexander, Lewis, and Janice C. Eberly. 2018. "Investment Hollowing Out." *IMF Economic Review* 66:5–30.

Alvarez, Fernando, and Robert Shimer. 2011. "Unions and Unemployment." Working Paper. http://home.uchicago.edu/~falvare/unions-print.pdf.

Alvarez, Fernando, and Marcelo Veracierto. 1999. "Labor-Market Policies in an Equilibrium Search Model." *NBER Macroeconomics Annual* 14:265–304.

Anderson, Eric, Sergio Rebelo, and Arlene Wong. 2019. "Markups across Space and Time." Working Paper 24434. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24434.

Arnold, David. 2020. "Mergers and Acquisitions, Local Labor Market Concentration, and Worker Outcomes." Job Market Paper. https://scholar.princeton.edu/sites/default/files/dharnold/files/jmp.pdf.

Auten, Gerald, and David Splinter. 2019. "Income Inequality in the United States: Using Tax Data to Measure Long-Term Trends." Working Paper. http://davidsplinter.com/AutenSplinter-Tax_Data_and_Inequality.pdf.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics* 135, no. 2: 645–709.

Autor, David, and Anna Salomons. 2018. "Is Automation Labor-Displacing? Productivity Growth, Employment, and the Labor Share." *Brookings Papers on Economic Activity*, Spring, 1–87.

Azar, José, Martin C. Schmalz, and Isabel Tecu. 2018. "Anticompetitive Effects of Common Ownership." *Journal of Finance* 73, no. 4: 1513–65.

Azar, José, and Xavier Vives. 2019. "Common Ownership and the Secular Stagnation Hypothesis." *American Economic Association Papers and Proceedings* 109:322–26.

Backus, Matthew, Christopher Conlon, and Michael Sinkinson. 2019. *The Common Ownership Hypothesis: Theory and Evidence*. Washington: Brookings Institution. https://www.brookings.edu/research/the-common-ownership-hypothesis-theory-and-explanation/.

Bakija, Jon, Adam Cole, and Bradley T. Heim. 2012. "Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from US Tax Return Data." Working Paper. https://web.williams.edu/Economics/wp/BakijaColeHeimJobsIncomeGrowthTopEarners.pdf.

Ball, Laurence, and N. Gregory Mankiw. 2002. "The NAIRU in Theory and Practice." *Journal of Economic Perspectives* 16, no. 4: 115–36.

Baqaee, David Rezza, and Emmanuel Farhi. 2020. "Productivity and Misallocation in General Equilibrium." *Quarterly Journal of Economics* 35, no. 1: 105–63.

Barkai, Simcha. Forthcoming. "Declining Labor and Capital Shares." *Journal of Finance*.

Barth, Erling, Alex Bryson, James C. Davis, and Richard Freeman. 2016. "It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States." *Journal of Labor Economics* 34, suppl. 2: S67–S97.

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.

Bell, Brian, Pawel Bukowski, and Stephen Machin. 2019. "Rent Sharing and Inclusive Growth." Working Paper 29. London: London School of Economics International Inequalities Institute. http://eprints.lse.ac.uk/101868/1/Bell_Bukowski_Machin_rent_sharing_and_inclusive_growth_wp_29.pdf.

Benmelech, Efraim, Nittai Bergman, and Hyunseob Kim. 2018. "Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages?" Working Paper 24307. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24307.

Bental, Benjamin, and Dominique Demougin. 2010. "Declining Labor Shares and Bargaining Power: An Institutional Explanation." *Journal of Macroeconomics* 32, no. 1: 443–56.

Bentolila, Samuel, and Gilles Saint-Paul. 2003. "Explaining Movements in the Labor Share." *BE Journal of Macroeconomics* 3, no. 1: 1–31.

Berger, David W., Kyle F. Herkenhoff, and Simon Mongey. 2019. "Labor Market Power." Working Paper 25719. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25719.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott. 2006. "Survival of the Best Fit: Exposure to Low-Wage Countries and the (Uneven) Growth of US Manufacturing Plants." *Journal of International Economics* 68, no. 1: 219–37.

Bernhardt, Annette, Rosemary Batt, Susan Houseman, and Eileen Appelbaum. 2016. "Domestic Outsourcing in the US: A Research Agenda to Assess Trends and Effects on Job Quality." Working Paper 102-16. Berkeley, Calif.: Institute for Research on Labor and Unemployment. https://irle.berkeley.edu/files/2016/Domestic-Outsourcing-in-the-US.pdf.

Berry, Steven T., Martin Gaynor, and Fiona Scott Morton. 2019. "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization." *Journal of Economic Perspectives* 33, no. 3: 44–68.

Bertola, Giuseppe, Francine D. Blau, and Lawrence M. Kahn. 2007. "Labor Market Institutions and Demographic Employment Patterns." *Journal of Population Economics* 20, no. 4: 833–67.

Bhuller, Manudeep, Andreas Kostøl, and Trond Vigtel. 2020. "How Broadband Internet Affects Labor Market Matching." Discussion Paper 12895. Bonn: Institute of Labor Economics. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3525222.

Bidwell, Matthew, Forrest Briscoe, Isabel Fernandez-Mateo, and Adina Sterling. 2013. "The Employment Relationship and Inequality: How and Why Changes in Employment Practices Are Reshaping Rewards in Organizations." *Academy of Management Annals* 7, no. 1: 61–121.

Bivens, Josh, and Lawrence Mishel. 2013. "The Pay of Corporate Executives and Financial Professionals as Evidence of Rents in Top 1 Percent Incomes." *Journal of Economic Perspectives* 27, no. 3: 57–78.

Bivens, Josh, Lawrence Mishel, and John Schmitt. 2018. *It's Not Just Monopoly and Monopsony: How Market Power Has Affected American Wages*. Washington: Economic Policy Institute. https://www.epi.org/publication/its-not-just-monopoly-and-monopsony-how-market-power-has-affected-american-wages/.

Blanchard, Olivier, Eugenio Cerutti, and Lawrence H. Summers. 2015. "Inflation and Activity—Two Explorations and Their Monetary Policy Implications." Working Paper w21726. Cambridge, Mass.: National Bureau of Economic Research.

Blanchard, Olivier J., and Francesco Giavazzi. 2003. "Macroeconomics of Regulation and Deregulation in Goods and Labor Markets." *Quarterly Journal of Economics* 118, no. 3: 879–907.

Blanchard, Olivier J., and Lawrence H. Summers. 1986. "Hysteresis and the European Unemployment Problem." *NBER Macroeconomics Annual* 1 (Fall): 15–78.

Blanchflower, David G., and Alex Bryson. 2004. "Union Relative Wage Effects in the United States and the United Kingdom." In *Proceedings of the Industrial*

Relations Research Association. San Diego: Industrial Relations Research Association.

Blanchflower, David G., Neil Millward, and Andrew J. Oswald. 1991. "Unionism and Employment Behaviour." *Economic Journal* 101, no. 407: 815–34.

Borjas, George J., and Valerie A. Ramey. 1995. "Foreign Competition, Market Power, and Wage Inequality." *Quarterly Journal of Economics* 110, no. 4: 1075–1110.

Bridgman, Benjamin. 2018. "Is Labor's Loss Capital's Gain? Gross Versus Net Labor Shares." *Macroeconomic Dynamics* 22, no. 8: 2070–87.

Bronfenbrenner, Kate. 2009. *No Holds Barred: The Intensification of Employer Opposition to Organizing*. Washington: Economic Policy Institute. https://www.epi.org/publication/bp235/.

Brown, Charles, and James Medoff. 1989. "The Employer Size Wage Effect." *Journal of Political Economy* 97, no. 5: 1027–59.

Brun, Lídia, and Ignacio González. 2017. "Tobin's Q and Inequality." Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3069980.

Bryson, Alex, and Morris M. Kleiner. 2019. "Re-examining Advances in Occupational Licensing Research: Issues and Policy Implications." *British Journal of Industrial Relations* 57, no. 4: 721–31.

Bulow, Jeremy I., and Lawrence H. Summers. 1986. "A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment." *Journal of Labor Economics* 4, no. 3, pt. 1: 376–414.

Caballero, Ricardo J., and Mohamad L. Hammour. 2005. "The Cost of Recessions Revisited: A Reverse-Liquidationist View." *Review of Economic Studies* 72, no. 2: 313–41.

Calmfors, Lars, and John Driffill. 1988. "Bargaining Structure, Corporatism and Macroeconomic Performance." *Economic Policy* 3, no. 6: 13–61.

Card, David. 1996. "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis." *Econometrica* 64, no. 4: 957–79.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics* 36, no. S1: S13–S70.

Cobb, J. Adam, and Ken-Hou Lin. 2017. "Growing Apart: The Changing Firm-Size Wage Premium, and Its Inequality Consequences." *Organization Science* 28, no. 3: 429–46.

Council of Economic Advisers (CEA). 2016. "Labor Market Monopsony: Trends, Consequences, and Policy Responses." Issue Brief. Washington: Council of Economic Advisers. https://obamawhitehouse.archives.gov/sites/default/files/page/files/20161025_monopsony_labor_mrkt_cea.pdf.

Covarrubias, Mattias, Germán Gutiérrez, and Thomas Philippon. 2019. "From Good to Bad Concentration? US Industries over the Past 30 Years." *NBER Macroeconomics Annual* 34:1–46.

Crouzet, Nicolas, and Janice C. Eberly. 2019. "Understanding Weak Capital Investment: The Role of Market Concentration and Intangibles." Working Paper 25869. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25869.

Crouzet, Nicolas, and Janice C. Eberly. 2020. "Rents and Intangible Capital: A 'Q+' Approach." Working Paper. https://www.kellogg.northwestern.edu/faculty/crouzet/html/papers/Qplus_latest.pdf.

Crump, Richard K., Stefano Eusepi, Marc Giannoni, and Ayşül Şahin. 2019. "A Unified Approach to Measuring u^*." *Brookings Papers on Economic Activity*, Spring, 143–238.

Dao, Mai Chi, Mitali Das, Zsoka Koczan, and Weicheng Lian. 2017. "Why Is Labor Receiving a Smaller Share of Global Income? Theory and Empirical Evidence." Working Paper 17/169. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2017/07/24/Why-Is-Labor-Receiving-a-Smaller-Share-of-Global-Income-Theory-and-Empirical-Evidence-45102.

Davis, Steven J., and John Haltiwanger. 1991. "Wage Dispersion between and within US Manufacturing Plants, 1963–86." *Brookings Papers on Economic Activity: Microeconomics*, 115–200.

Davis, Steven J., and John Haltiwanger. 1996. "Employer Size and the Wage Structure in US Manufacturing." *Annales D'Economie et de Statistique* 41/42:323–67.

Davis, Steven J., John C. Haltiwanger, Kyle Handley, Ben Lipsius, Josh Lerner, and Javier Miranda. 2019. "The Economic Effects of Private Equity Buyouts." Working Paper 26371. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26371.

Davis, Steven J., John Haltiwanger, Ron Jarmin, and Javier Miranda. 2006. "Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms." *NBER Macroeconomics Annual* 21:107–80.

Davis, Steven J., and Till von Wachter. 2011. "Recessions and the Cost of Job Loss." *Brookings Papers on Economic Activity*, Fall, 1–72.

De Loecker, Jan, and Jan Eeckhout. 2019. "Global Market Power." Working Paper. http://www.janeeckhout.com/wp-content/uploads/Global.pdf.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics* 135, no. 2: 561–644.

Denice, Patrick, and Jake Rosenfeld. 2018. "Unions and Nonunion Pay in the United States, 1977–2015." *Sociological Science* 5:541–61.

Dickens, William T., and Lawrence F. Katz. 1987. "Interindustry Wage Differences and Industry Characteristics." In *Unemployment and the Structure of Labor Markets*, edited by Kevin Lang and Jonathan S. Leonard. New York: Basil Blackwell.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach." *Econometrica* 64, no. 5: 1001–44.

Dorn, David, Johannes F. Schmieder, and James R. Spletzer. 2018. "Domestic Outsourcing in the United States." Working Paper. https://www.dol.gov/sites/dolgov/files/OASP/legacy/files/Domestic-Outsourcing-in-the-United-States.pdf.

Dube, Arindrajit, and Ethan Kaplan. 2010. "Does Outsourcing Reduce Wages in the Low-Wage Service Occupations? Evidence from Janitors and Guards." *Industrial and Labor Relations Review* 63, no. 2: 287–306.

Ebell, Monique, and Christian Haefke. 2009. "Product Market Deregulation and the US Employment Miracle." *Review of Economic Dynamics* 12, no. 3: 479–504.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2018. "How Costly Are Markups?" Working Paper 24800. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24800.

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.: https://www.nber.org/papers/w24287.

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin. 2013. "The Decline of the US Labor Share." *Brookings Papers on Economic Activity*, Fall, 1–63.

Erickson, Christopher L., and Daniel J. B. Mitchell. 2007. "Monopsony as a Metaphor for the Emerging Post-Union Labour Market." *International Labour Review* 146, nos. 3/4: 163–87.

Even, William E., and David A. Macpherson. 2012. "Is Bigger Still Better? The Decline of the Wage Premium at Large Firms." *Southern Economic Journal* 78, no. 4: 1181–1201.

Farber, Henry S. 2005. "Nonunion Wage Rates and the Threat of Unionization." *Industrial and Labor Relations Review* 58, no. 3: 335–52.

Farber, Henry S., Daniel Herbst, Ilyana Kuziemko, and Suresh Naidu. 2018. "Unions and Inequality over the Twentieth Century: New Evidence from Survey Data." Working Paper 24587. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24587.

Farhi, Emmanuel, and François Gourio. 2018. "Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia." *Brookings Papers on Economic Activity*, Fall, 147–250.

Figura, Andrew, and David Ratner. 2015. "The Labor Share of Income and Equilibrium Unemployment." FEDS Notes. Washington: Board of Governors of the Federal Reserve System.

Fortin, Nicole, Thomas Lemieux, and Neil Lloyd. 2018. "Labor Market Institutions and the Distribution of Wages: The Role of Spillover Effects." Ann Arbor: University of Michigan.

Freeman, Richard B., and James L. Medoff. 1984. *What Do Unions Do?* New York: Basic Books.

Furman, Jason, and Alan Krueger. 2016. "Why Aren't Workers Getting Raises? Blame the Monopsony." *Wall Street Journal*, November 3.

Galbraith, John Kenneth. 1952. *American Capitalism: The Concept of Countervailing Power*. Boston: Houghton Mifflin.

Galí, Jordí. 2020. "Insider-Outsider Labor Markets, Hysteresis and Monetary Policy." Working Paper 1506. Barcelona: Universitat Pompeu Fabra. https://econ-papers.upf.edu/papers/1506.pdf.

Ganapati, Sharat. Forthcoming. "Growing Oligopolies, Prices, Output, and Productivity." *American Economic Journal: Microeconomics*.

Geroski, Paul, Paul Gregg, and John Van Reenen. 1995. "Market Imperfections and Employment." Working Paper 5. Paris: Organization for Economic Cooperation and Development. https://www.oecd-ilibrary.org/employment/market-imperfections-and-employment_824728761027.

Gibbons, Robert, and Lawrence Katz. 1992. "Does Unmeasured Ability Explain Inter-Industry Wage Differentials?" *Review of Economic Studies* 59, no. 3: 515–35.

Gittleman, Maury, Mark A. Klee, and Morris M. Kleiner. 2018. "Analyzing the Labor Market Outcomes of Occupational Licensing." *Industrial Relations: A Journal of Economy and Society* 57, no. 1: 57–100.

Glover, Andrew, and Jacob Short. 2018. "Demographic Origins of the Decline in Labor's Share." Working Paper. http://andyecon.weebly.com/uploads/5/1/3/8/5138005/ls_age_mar2018.pdf.

González, Ignacio, and Pedro Trivín. 2019. "The Global Rise of Asset Prices and the Decline of the Labor Share." Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2964329.

Greenwald, Daniel L., Martin Lettau, and Sydney C. Ludvigson. 2019. "How the Wealth Was Won: Factors Shares as Market Fundamentals." Working Paper 25769. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25769.

Groshen, Erica L. 1991. "Source of Intra-Industry Wage Dispersion: How Much Do Employers Matter?" *Quarterly Journal of Economics* 106, no. 3: 869–84.

Grout, Paul A. 1984. "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach." *Econometrica* 52, no. 2: 449–60.

Grullon, Gustavo, Yelena Larkin, and Roni Michaely. 2019. "Are US industries Becoming More Concentrated?" *Review of Finance* 23, no. 4: 697–743.

Gutiérrez, Germán, and Thomas Philippon. 2017. "Investmentless Growth: An Empirical Investigation." *Brookings Papers on Economic Activity*, Fall, 89–190.

Gutiérrez, Germán, and Thomas Philippon. 2019. "Fading Stars." In *American Economic Association Papers and Proceedings*, vol. 109. Pittsburgh, Pa.: AEA Publications.

Gutiérrez, Germán, and Sophie Piton. Forthcoming. "Revisiting the Global Decline of the (Non-housing) Labor Share." *American Economic Review: Insights*.

Hall, Robert E. 1975. "The Rigidity of Wages and the Persistence of Unemployment." *Brookings Papers on Economic Activity*, no. 2: 300–349.

Hall, Robert E. 2018. "Using Empirical Marginal Cost to Measure Market Power in the US Economy." Working Paper 18118. Stanford, Calif.: Hoover Institution. https://www.hoover.org/sites/default/files/research/docs/18118-hall.pdf.

Haltiwanger, John C., and James R. Spletzer. 2020. "Between Firm Changes in Earnings Inequality: The Dominant Role of Industry Effects." Working Paper 26786. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26786.

Handwerker, Elizabeth Weber. 2018. "Increased Concentration of Occupations, Outsourcing, and Growing Wage Inequality in the United States." Working Paper 522. Washington: US Bureau of Labor Statistics. https://www.bls.gov/osmr/research-papers/2020/pdf/ec200030.pdf.

Hazell, Jonathon. 2019. "The Falling Labor Share Dampens Unemployment Fluctuations." Working Paper. https://economics.mit.edu/files/18831.

Hollister, Matissa N. 2004. "Does Firm Size Matter Anymore? The New Economy and Firm Size Wage Effects." *American Sociological Review* 69, no. 5: 659–76.

Holzer, Harry J., Lawrence F. Katz, and Alan B. Krueger. 1991. "Job Queues and Wages." *Quarterly Journal of Economics* 106, no. 3: 739–68.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. 1993. "Earnings Losses of Displaced Workers." *American Economic Review* 83, no. 4: 685–709.

Jaumotte, Florence, and Carolina Osorio Buitron. 2015. "Inequality and Labor Market Institutions." Staff Discussion Note SDN/15/14. Washington: International Monetary Fund. https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2016/12/31/Inequality-and-Labor-Market-Institutions-42987.

Karabarbounis, Loukas, and Brent Neiman. 2014. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129, no. 1: 61–103.

Karabarbounis, Loukas, and Brent Neiman. 2018. "Accounting for Factorless Income." *NBER Macroeconomics Annual* 33:167–228.

Katz, Lawrence F., and Alan Krueger. 1999. "New Trend in Unemployment?" *Brookings Review* 17, no. 4: 4–8.

Katz, Lawrence F., and Lawrence H. Summers. 1989. "Industry Rents: Evidence and Implications." *Brookings Papers on Economic Activity: Microeconomics*, 209–90.

Kehrig, Matthias, and Nicolas Vincent. 2020. "The Micro-Level Anatomy of the Labor Share Decline." Working Paper 25275. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25275.

Kim, Changhwan, and Arthur Sakamoto. 2008. "Declining Inter-Industry Wage Dispersion in the US" *Social Science Research* 37, no. 4: 1081–1101.

Kleiner, Morris M., and Alan B. Krueger. 2013. "Analyzing the Extent and Influence of Occupational Licensing on the Labor Market." *Journal of Labor Economics* 31, no. 2: S173–S202.

Kleiner, Morris M., and Evan J. Soltas. 2019. "A Welfare Analysis of Occupational Licensing in US States." Working Paper w26383. Cambridge, Mass.: National Bureau of Economic Research.

Kohn, Donald L. 2005. "Globalization, Inflation, and Monetary Policy." Remarks at the James R. Wilson Lecture Series, College of Wooster, Wooster, Ohio, October 11.

Kristal, Tali. 2010. "Good Times, Bad Times: Postwar Labor's Share of National Income in Capitalist Democracies." *American Sociological Review* 75, no. 5: 729–63.

Krueger, Alan B. 2018. “Reflections on Dwindling Worker Bargaining Power and Monetary Policy.” Speech given at the Jackson Hole Economic Symposium, August 24.

Krueger, Alan B., and Orley Ashenfelter. 2018. “Theory and Evidence on Employer Collusion in the Franchise Sector.” Working Paper 24831. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24831.

Krueger, Alan B., and Lawrence H. Summers. 1988. “Efficiency Wages and the Inter-Industry Wage Structure.” *Econometrica* 56, no. 2: 259–93.

Kuhn, Peter, and Hani Mansour. 2014. “Is Internet Job Search Still Ineffective?” *Economic Journal* 124, no. 581: 1213–33.

Lachowska, Marta, Alexandre Mas, Rafaelle D. Saggio, and Stephen A. Woodbury. 2020. “Do Firm Effects Drift? Evidence from Washington Administrative Data.” Working Paper 26653. Cambridge, Mass.: National Bureau of Economic Research.

Lachowska, Marta, Alexandre Mas, and Stephen A. Woodbury. 2018. “Sources of Displaced Workers’ Long-Term Earnings Losses.” Working Paper 24217. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24217.

Leduc, Sylvain, and Daniel J. Wilson. 2017. “Has the Wage Phillips Curve Gone Dormant?” Economic Letter 2017-30. Federal Reserve Bank of San Francisco. https://www.frbsf.org/economic-research/files/el2017-30.pdf.

Lee, David S., and Alexandre Mas. 2012. “Long-Run Impacts of Unions on Firms: New Evidence from Financial Markets 1961–1999.” *Quarterly Journal of Economics* 127, no. 1: 333–78.

Leicht, Kevin T. 1989. “On the Estimation of Union Threat Effects.” *American Sociological Review* 54, no. 6: 1035–47.

Leonard, Jonathan S. 1992. “Unions and Employment Growth.” *Industrial Relations* 31, no. 1: 80–94.

Levine, David I., Dale Belman, Gary Charness, Erica L. Groshen, and K. C. O’Shaughnessy. 2002. “How New Is the ‘New Employment Contract’?: Evidence from North American Pay Practices.” Kalamazoo, Mich.: Upjohn Institute for Employment Research.

Levy, Frank, and Peter Temin. 2007. “Inequality and Institutions in 20th Century America.” Working Paper 13106. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w13106.

MacDonald, Ian M., and Robert M. Solow. 1981. “Wage Bargaining and Employment.” *American Economic Review* 71, no. 5: 896–908.

Manning, Alan. 1990. “Imperfect Competition, Multiple Equilibria and Unemployment Policy.” *Economic Journal* 100, no. 400: 151–62.

Manning, Alan. 2003. *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton, N.J.: Princeton University Press.

McNicholas, Celine, Margaret Poydock, Julia Wolfe, Ben Zipperer, Gordon Lafer, and Lola Loustaunau. 2019. *Unlawful: US Employers Are Charged with Violating Federal Law in 41.5% of All Union Election Campaigns*. Washington: Economic Policy Institute.

Mishel, Lawrence. 2018. *Yes, Manufacturing Still Provides a Pay Advantage, but Staffing Firm Outsourcing Is Eroding It*. Washington: Economic Policy Institute. https://www.epi.org/publication/manufacturing-still-provides-a-pay-advantage-but-outsourcing-is-eroding-it/.

Mishel, Lawrence, Josh Bivens, Elise Gould, and Heidi Shierholz. 2012. *The State of Working America*, 12th ed. Ithaca, N.Y.: Cornell University Press.

Molloy, Raven, Riccardo Trezzi, Christopher L. Smith, and Abigail Wozniak. 2016. "Understanding Declining Fluidity in the US Labor Market." *Brookings Papers on Economic Activity*, Spring, 183–259.

Montgomery, Edward. 1989. "Employment and Unemployment Effects of Unions." *Journal of Labor Economics* 7, no. 2: 170–90.

Mortensen, Dale T., and Christopher A. Pissarides. 1999. "Job Reallocation, Employment Fluctuations and Unemployment." In *Handbook of Macroeconomics*, vol. 1B. Amsterdam: North Holland.

Neumark, David, and Michael L. Wachter. 1995. "Union Effects on Nonunion Wages: Evidence from Panel Data on Industries and Cities." *Industrial and Labor Relations Review* 49, no. 1: 20–38.

Oswald, Andrew J. 1982. "Trade Unions, Wages and Unemployment: What Can Simple Models Tell Us?" *Oxford Economic Papers* 34, no. 3: 526–45.

Paul, Sanjukta M. 2016. "The Enduring Ambiguities of Antitrust Liability for Worker Collective Action." *Loyola University Chicago Law Journal* 47:969–1048.

Peltzman, Sam. 2018. "Productivity and Prices in Manufacturing during an Era of Rising Concentration." Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3168877.

Peoples, James. 1998. "Deregulation and the Labor Market." *Journal of Economic Perspectives* 12, no. 3: 111–30.

Phelps, Edmund S. 1968. "Money-Wage Dynamics and Labor-Market Equilibrium." *Journal of Political Economy* 76, no. 4, pt. 2: 678–711.

Philippon, Thomas. 2020. *The Great Reversal: How America Gave Up on Free Markets*. Cambridge, Mass.: Harvard University Press.

Philippon, Thomas, and Ariell Reshef. 2012. "Wages and Human Capital in the US Finance Industry: 1909–2006." *Quarterly Journal of Economics* 127, no. 4: 1551–609.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2018. "Distributional National Accounts: Methods and Estimates for the United States." *Quarterly Journal of Economics* 133, no. 2: 553–609.

Ratner, David, and Jae Sim. 2020. "Who Killed the Phillips Curve? A Murder Mystery." Working Paper. https://www.researchgate.net/publication/339384053_Who_Killed_the_Phillips_Curve_A_Murder_Mystery.

Rinz, Kevin. 2018. "Labor Market Concentration, Earnings Inequality, and Earnings Mobility." Working Paper 2018-10. Washington: US Census Bureau, Center for Administrative Research and Applications. https://www.census.gov/library/working-papers/2018/adrm/carra-wp-2018-10.html.

Rosenfeld, Jake. 2014. *What Unions No Longer Do*. Cambridge, Mass.: Harvard University Press.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Nicholas Trachter. Forthcoming. "Diverging Trends in National and Local Concentration." *NBER Macroeconomics Annual* 35.

Salinger, Michael A. 1984. "Tobin's Q, Unionization, and the Concentration-Profits Relationship." *RAND Journal of Economics* 15, no. 2: 159–70.

Schmalz, Martin C. 2018. "Common-Ownership Concentration and Corporate Conduct." *Annual Review of Financial Economics* 10:413–48.

Schmitt, John, and Alexandra Mitukiewicz. 2012. "Politics Matter: Changes in Unionisation Rates in Rich Countries, 1960–2010." *Industrial Relations Journal* 43, no. 3: 260–80.

Schubert, Gregor, Anna Stansbury, and Bledi Taska. 2020. "Monopsony and Outside Options." Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3599454.

Shapiro, Carl. 2018. "Antitrust in a Time of Populism." *International Journal of Industrial Organization* 61:714–48.

Shimer, Robert. 1998. "Why Is the US Unemployment Rate So Much Lower?" *NBER Macroeconomics Annual* 13:11–61.

Shleifer, Andre, and Lawrence H. Summers. 1988. "Breach of Trust in Hostile Takeovers." In *Corporate Takeovers: Causes and Consequences*, edited by Alan J. Auerbach. Chicago: University of Chicago Press.

Slichter, Sumner H. 1950. "Notes on the Structure of Wages." *Review of Economics and Statistics* 32, no. 1: 80–91.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. "The Rise of Pass-Throughs and the Decline of the Labor Share." Working Paper. http://ericzwick.com/labshare/labshare.pdf.

Sokolova, Anna, and Todd Sorensen. 2020. "Monopsony in Labor Markets: A Meta-Analysis." Working Paper. Washington: Washington Center for Equitable Growth. https://equitablegrowth.org/working-papers/monopsony-in-labor-markets-a-meta-analysis/.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2019. "Firming Up Inequality." *Quarterly Journal of Economics* 134, no. 1: 1–50.

Sorkin, Isaac. 2018. "Ranking Firms Using Revealed Preference." *Quarterly Journal of Economics* 133, no. 3: 1331–93.

Stansbury, Anna, and Lawrence H. Summers. 2019. "Productivity and Pay: Is the Link Broken?" In *Facing Up to Low Productivity Growth*, edited by Adam S. Posen and Jeromin Zettelmeyer. Washington: Peterson Institute for International Economics.

Starr, Evan, J. J. Prescott, and Norman Bishara. 2019. "Noncompetes in the US Labor Force." Working Paper 18-013. Ann Arbor, Mich.: University of Michigan Law and Economics. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2625714.

Steinbaum, Marshall. 2019. "Antitrust, the Gig Economy, and Labor Market Power." *Law and Contemporary Problems* 82:45–64.

Stevenson, Betsey. 2009. "The Internet and Job Search." In *Studies of Labor Market Intermediation*, edited by David H. Autor. Chicago: University of Chicago Press.

Summers, Lawrence H. 1986. "Why Is the Unemployment Rate So Very High Near Full Employment?" *Brookings Papers on Economic Activity*, no. 2: 339–96.

Summers, Lawrence H. 1988. "Relative Wages, Efficiency Wages, and Keynesian Unemployment." Working Paper w2590. Cambridge, Mass.: National Bureau of Economic Research.

Summers, Lawrence H. 2013. Speech given at the IMF Fourteenth Annual Research Conference, Washington, D.C., November 8.

Syverson, Chad. 2019. "Macroeconomics and Market Power: Context, Implications, and Open Questions." *Journal of Economic Perspectives* 33, no. 3: 23–43.

Traina, James. 2018. "Is Aggregate Market Power Increasing? Production Trends Using Financial Statements." Working Paper 17. Chicago: University of Chicago, Stigler Center for the Study of the Economy and the State. https://research.chicagobooth.edu/-/media/research/stigler/pdfs/workingpapers/17isaggregatemarketpowerincreasing.pdf.

Tüzemen, Didem. 2019. "Job Polarization and the Natural Rate of Unemployment in the United States." *Economics Letters* 175:97–100.

Ulman, Lloyd. 1965. "Labor Mobility and the Industrial Wage Structure in the Postwar United States." *Quarterly Journal of Economics* 79, no. 1: 73–97.

Van Reenen, John. 2018. "Increasing Differences between Firms: Market Power and the Macro-economy." In *Economic Policy Symposium Proceedings: Changing Market Structures and Implications for Monetary Policy.* Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Webber, Douglas A. 2018. "Employment Adjustment over the Business Cycle: The Impact of Competition in the Labor Market." Discussion Paper 11887. IZA Institute of Labor Economics.

Weil, David. 2014. *The Fissured Workplace: Why Work Became So Bad for So Many and What Can Be Done to Improve It*. Cambridge, Mass.: Harvard University Press.

Weil, David. 2019. "Understanding the Present and Future of Work in the Fissured Workplace Context." *Russell Sage Foundation Journal of the Social Sciences* 5, no. 5: 147–65.

Western, Bruce, and Jake Rosenfeld. 2011. "Unions, Norms, and the Rise in US Wage Inequality." *American Sociological Review* 76, no. 4: 513–37.

Wilmers, Nathan. 2018. "Wage Stagnation and Buyer Power: How Buyer-Supplier Relations Affect US Workers' Wages, 1978 to 2014." *American Sociological Review* 83, no. 2: 213–42.

Yellen, Janet. 1984. "Efficiency Wage Models of Unemployment." *American Economic Review* 74, no. 2: 200–205.

Comments and Discussion

COMMENT BY

STEVEN J. DAVIS Anna Stansbury and Lawrence Summers advance the thesis that worker power fell in recent decades, leading to a loss of labor rents. As explanations for the fall in worker power, they stress two factors: first, institutional forces such as the shrinking role of private sector unions and second, increases in shareholder power and shareholder activism that drove firms to cut wages, outsource labor, and rely more heavily on contract workers.

The authors cover a lot of ground, and I will remark only on parts of their study. They make a strong case that labor rent premiums fell for many American workers. They also show that falling worker power and labor rents can potentially explain major developments that other authors have attributed to rising product market power or rising monopsony power—including labor's falling share of national income, the rise of measured markups, increased market valuations of publicly traded firms, and a fall in the non-accelerating inflation rate of unemployment (NAIRU). These are important contributions.

Less persuasive are their headline claims about the magnitude of falling labor rents, especially when understood as a redistribution of rents from labor to capital. I make several points in this regard and conclude that we do not yet have a confident quantification of what happened to aggregate labor rents (as a share of value added) in recent decades. I also offer some remarks on the role of technological developments and globalization in driving declines in unionization and the average rent premium.

THE MAGNITUDE OF FALLING LABOR RENTS Stansbury and Summers estimate that labor rents fell "from 12 percent of net value added in the nonfinancial corporate business sector in the early 1980s to 6 percent in the 2010s." Figure 5 plots the estimated labor rent share from 1982 to 2016, and table 1 reports the separate contributions of union rents, firm size rents, and industry rents.

These estimates are key inputs into their later analyses. They derive them in two steps. First, they use Current Population Survey (CPS) micro data to estimate average log wage premia associated with union status, employer size, and industry (conditional on various controls). For each category of rents, they designate a reference sector or group with zero rents by assumption, and they interpret one-half of the log wage premium relative to the reference sector or group as a rent premium. Second, they combine these sectoral and group-level rent premium estimates with data on the share of compensation in the nonfinancial corporate sector, the industry distribution of compensation, the union coverage rate, and the distribution of compensation by firm size as spelled out by the formulas in section II. My remarks below pertain to their first step.

At the outset, it's useful to distinguish between two empirical objects: (a) the average rent premium among workers in a given sector or group, and (b) the natural log of labor rents divided by (compensation minus labor rents) in the sector or group. Stansbury and Summers estimate (a) by fitting equal-weighted log wage regressions to CPS micro data by year. When they plug their sectoral and group rent premium estimates into their formulas, they implicitly equate (a) to (b). However, these two objects have the same value only under special circumstances—for example, when all workers in the sector or group receive the same rent premium. Likewise, changes over time in (a) and (b) are identical only under special circumstances.

Of course, objects (a) and (b) are roughly the same under a broader range of circumstances. So, the issue is whether (a) and (b) are roughly the same—and changed by roughly the same amount over time—in the circumstances that unfolded in the US economy in recent decades.

There are good reasons to think not. First, Stansbury and Summers show in online appendix figure B.4 that the rent premium fell almost twice as much from 1984 to 2016 for workers with a noncollege education as for those with a college education. The rent premium was about 15–16 percent for both groups in 1984. Because they earn more, each college-educated worker contributes more to total labor rents than each noncollege worker. Likewise, each college-educated worker properly gets a larger weight in quantifying the fall in labor rents over time. In contrast, the approach taken by Stansbury and Summers weights workers equally in the quantification of labor rents and their changes over time. This aspect of their approach overstates the fall in labor rents—that is, object (b)—when high-wage workers experience smaller declines in rent premia.

Second, online appendix figure B.4 also says that college-educated workers have enjoyed a higher rent premium than noncollege workers since the early 1990s. Because the college versus noncollege wage gap has expanded over time, the appropriate weight on each college-educated worker in the calculation of total labor rents is increasing over time.[1] But, as already noted, Stansbury and Summers maintain equal weights across workers. Equal weighting overstates the fall in object (b) when workers with higher rent premia experience more rapid growth in the non-rent component of their compensation. That pattern has unfolded in recent decades for college-educated workers relative to noncollege workers.

Third, other studies point to rising rent premia in recent decades for senior executives, managers, and highly compensated professionals—especially in the financial sector. Stansbury and Summers take note of several such studies in section V.C. These other studies, when combined with the authors' evidence that rent premia fell in recent decades for the average worker, strongly suggest that labor rents were redistributed from the bottom and middle of the wage distribution to the top of the wage distribution. In these circumstances, the average rent premium can fall substantially even when labor rents as a share of total compensation are unchanged. That is, object (a) aggregated over sectors and groups can fall substantially, even when object (b) (also aggregated) remains unchanged. More generally, a redistribution of labor rents from the majority of workers to those at the top end yields a fall in (a) relative to (b).

Stansbury and Summers acknowledge in section V.A that some of what they measure as lost labor rents could instead be a redistribution of rents from the majority of workers to top management, executives, and highly compensated workers in the financial sector. They go on to argue in online appendix B.2 that the "degree to which the exclusion of top-earning workers in finance might affect our calculations is relatively limited." But the calculation they offer to support that claim involves a counterfactual change in the equal-weighted average rent premium—object (a)—whereas the impact on object (b) turns on how the exclusion of top earners affects the pay-weighted average rent premium. Since highly compensated financial professionals earn much more than average workers, the evolution of their rent premium over time has a much bigger impact on (b) than on (a). The same point applies to the labor rents of top managers, executives, and other highly compensated persons.

1. The college versus noncollege wage gap has increased more than the college versus noncollege gap in rent premiums. Thus, the non-rent part of wages has risen for college-educated workers relative to noncollege workers.

In principle, one can recover estimates of object (b) by fitting pay-weighted regressions to the CPS micro data instead of the equal-weighted regressions that underlie figure 5 and table 1 in their paper. Stansbury and Summers report selected results using wage-weighted regressions in their online appendix figures A.7 and A.8. Comparing figure A.7 and figure 3 shows that the cross-industry variance of labor rents evolves similarly over time whether estimated from a wage-weighted regression or an equal-weighted regression. This comparison suggests that objects (a) and (b) moved similarly over time for industry rent premia. Unfortunately, the comparison is distorted by serious weaknesses in the CPS micro data.

That brings me to the limitations of CPS data as a tool for quantifying total labor rents and their evolution over time. As Stansbury and Summers note, CPS data are top-coded and nonresponse rates are high for persons in the top tail of the earnings distribution. That may not matter much for estimating object (a), but it potentially matters a great deal for estimating object (b) because persons in the upper part of the earnings distribution account for a large and growing share of overall labor compensation in recent decades.

To appreciate the dimensions of the issue, consider some particulars of top coding and nonresponse in the CPS. Stansbury and Summers report that the share of wage earners with top-coded earnings in their sample varies from 1 percent to 5 percent, depending on the year. They also report that the share of workers with top-coded earnings in the finance, insurance, and real estate (FIRE) sector rose from 2 percent in 2000 to 9 percent by 2019. Based on US tax records, Smith, Zidar, and Zwick (2020, fig. 3.G) report that the top 1 percent of wage earners accounted for about 15 percent of aggregate wages in 2016, roughly double its share in 1980. The top 10 percent of wage earners accounted for about 42 percent of aggregate wages in 2016, up by about 10 percentage points since 1980. These statistics underscore the scope for rising labor rents at the upper end of the earnings distribution to go undetected in the CPS.

Philippon and Reshef (2012) combine CPS micro data for the bottom 90 percent of the wage distribution with BEA data on total compensation by sector to back out the average wage for top-decile earners by sector. They find that the average wage in the top decile of finance went from parity with the top-decile average wage in the nonfarm private economy in 1980 to a premium of more than 80 percent in 2010. Other evidence in Philippon and Reshef (2012) indicates that much of the rise in relative wages of top-decile finance workers reflects an increase in their rent premium. They also estimate that the finance sector accounts for 6–25 percent of the overall increase in US wage inequality since 1980, with larger percentages for measures that give more weight to inequality in the upper parts of the distribution.

Nonresponse is another key feature of the CPS (and other household surveys) that hampers estimation of labor rent premia, especially in the upper parts of the earnings distribution. Unit nonresponse rates in the CPS-ASEC range from 16 to 20 percent in the period from 1997 to 2011 (Meyer, Mok, and Sullivan 2015, fig. 1). Item nonresponse rates on earnings questions in the CPS-ASEC rose from about 9 percent in 1987 to about 24 percent in 2015 (Bollinger and others 2019, fig. 1). The total earnings nonresponse rate in CPS-ASEC data—encompassing those who decline to participate in the ASEC and those who participate but fail to answer the earnings questions—rose from about 18 percent in 1987 to 43 percent in 2015. Item nonresponse rates to earnings questions in the CPS-ORG data exceed 35 percent in recent years (Bollinger and others 2019).

Working with CPS-ASEC data for 2006–2011 linked to Social Security earnings records, Bollinger and others (2019) examine item nonresponse rates across the whole distribution of earnings as measured in the Social Security records. They find a *U*-shaped nonresponse pattern, which they characterize as "trouble in the tails." Among full-time full-year workers, they find 30 percent nonresponse rates to CPS earnings questions at the top end of the earnings distribution. Moreover, nonresponse rates continue to exhibit a *U*-shaped relationship to earnings after conditioning on a rich set of controls for demographic characteristics and employment status. Put differently, they reject the hypothesis that nonresponse is random, and they still reject it when controlling for the types of observables available in the CPS.

These facts about high, rising, and nonrandom nonresponse rates in the CPS micro data raise concerns even about the Stansbury and Summers estimates of empirical object (a).[2] These facts about nonresponse rates and my earlier remarks about top coding lead me to the conclusion that CPS data do not provide a sound basis for estimating empirical object (b) and its evolution in recent decades.

To summarize, I see the evidence provided by Stansbury and Summers as showing that the equal-weighted average rent premium fell for private sector American workers in recent decades.[3] I see the question of what

2. Quite sensibly, Stansbury and Summers do not use imputed CPS earnings data.

3. In contrast to the fall in the average rent premium among private sector workers, the average rent premium among public sector employees has increased substantially since the early 1980s (Gittleman and Pierce 2011).

happened to aggregate labor rents as largely open. Moreover, I do not think that CPS micro data are adequate for developing a persuasive analysis of this second question, that is, for how object (b) has moved in recent decades.

To be sure, questions about the evolution of rent premiums for the average worker (or the majority of workers) are interesting and important. But answering questions about changes in rent premiums for the average worker is not sufficient to determine what happened to total labor rents or labor rents as a share of value added. Nor is it sufficient to discern whether, and how much, the loss of rents for the average worker reflects a redistribution of rents to capital owners.

ON THE ROLE OF TECHNOLOGICAL DEVELOPMENT AND GLOBALIZATION What drove the decline in rent premia for the majority of American workers in recent decades? Here, I see a greater role for globalization and technological changes than Stansbury and Summers. I will briefly sketch some reasons why, but my remarks only scratch the surface of a complex set of issues that warrants more research.

Consider developments in the US manufacturing sector. Historically, manufacturing workers had high unionization rates and earned high wages compared to observationally similar workers in other sectors. As of 1977, the union membership rate was 35.5 percent in manufacturing and 17.6 percent in the nonmanufacturing part of the nonfarm private sector. The corresponding coverage rates were 37.6 percent and 19.2 percent.[4] Unionization rates were higher yet among manufacturing production workers. Tabulations in Freeman (1980, table 1) using CPS data from 1973 to 1975 imply that 55 percent of blue-collar male workers in the manufacturing sector were union members. Alternatively, under the assumption that nonproduction workers in manufacturing had the same unionization rate as the nonmanufacturing part of the private sector, the implied membership rate for manufacturing production workers is 43.2 percent.

These high-wage, heavily unionized jobs became a steadily shrinking share of aggregate employment over time, largely because of automation, foreign outsourcing, and greater competition from foreign producers. Production workers in the manufacturing sector fell from 16.3 percent of all

4. Membership and coverage rates are from Hirsch and Macpherson (2003), as updated at unionstats.com. I back out the rates in the nonmanufacturing part of the nonfarm private sector using the fact that manufacturing accounted for 25.1 percent of nonfarm private sector employees in 1977.

nonfarm employees in 1977 to 6 percent in 2016.[5] Now consider a counterfactual in which 10.3 percent of nonfarm employees shift from production jobs in manufacturing to nonmanufacturing jobs while the unionization rate stays unchanged for each category of jobs. This counterfactual yields a drop in the private sector unionization rate of 2.6 to 3.8 percentage points, which amounts to 17–25 percent of the overall 1977–2016 drop in the union membership rate among nonfarm private sector employees. This counterfactual suggests that globalization and automation played significant roles in shrinking the private sector unionization rate in recent decades. By design, the counterfactual speaks only to the potential effects of globalization and automation working through the share of overall employment accounted for by manufacturing production workers. These same forces may also affect unionization and rent premia through other channels.

Stansbury and Summers remark that stronger foreign competition may have eroded the market power of US manufacturers in recent decades. If correct, this characterization points to another channel through which globalization potentially drove a shrinking unionization rate: when there are fewer monopoly profits to share, workers have less to gain by opting for unions that exist partly to extract monopoly profits. In these circumstances, it becomes more challenging for unions to win the certification elections that grant collective bargaining rights, and it becomes less attractive for national union organizations to invest in certification elections. In this connection, note that unionization fell at a faster rate from 1984 to 2019 in the manufacturing sector than in any other industry sector except for mining. See online appendix figure C.2 in Stansbury and Summers. The relatively rapid fall of unionization within manufacturing—from a high initial level—reinforces the view that globalization was a significant factor behind falling unionization.

Stansbury and Summers also remark that unions can lead to rents for workers at nonunion employers through threat effects. If union threat effects outweigh the countervailing effects of unionism on pay in nonunion jobs, then any external force that causes a decline in unionization leads to falling rent premia for nonunion workers. Those external forces may be policy-oriented or institutional in nature, or they may reflect other forces

5. I computed these statistics by combining data on manufacturing and nonfarm employment from the Current Employment Statistics with data on the production worker share of manufacturing employment from the Annual Survey of Manufactures. The corresponding figures for all manufacturing workers (inclusive of nonproduction workers) are 22 percent of nonfarm employees in 1977 and 8.6 percent in 2016.

such as greater foreign competition. In other words, threat effects amplify the impact of declining unionization on labor rents regardless of what drives the decline.

Finally, it's worth remarking that—for any given unionization rate—stronger foreign competition is likely to erode the rent premium among union and nonunion employees of affected firms. The reason is simple: when employers have lower profits, there is less to share with workers in the form of labor rents. This is yet another channel through which stronger foreign competition lowers the average labor rent premium.

In short, my remarks suggest that automation and foreign competition reduced the average rent premium among American workers in the private sector by (1) lowering the share of employment accounted for by manufacturing production workers, a heavily unionized group that had enjoyed high rent premiums; (2) making unionization less attractive within the manufacturing sector; (3) lowering rent premiums among nonunion workers through diminished union threat effects; and (4) reducing the profitability of firms facing more intense foreign competition.

I turn now to another development that may contribute to the fall in the average labor rent premium: advances in employee monitoring technologies and their deployment in the workplace. An important class of efficiency-wage models attributes labor rents to the difficulties that employers face in monitoring worker performance. In these models, improvements in the ability of employers to detect subpar worker effort (shirking) leads to a fall in the equilibrium rent premium. See, for example, the one-sector model of Shapiro and Stiglitz (1984) and the multisector model of Bulow and Summers (1986).

Technologies for tracking vehicles and workers have become common in trucking, delivery services, and field service operations in recent decades (Dutta 2012). Tracking covers vehicle location, speed, idle time, fuel consumption, customer contact, delivery items, and more. The web and social media apps have also made it easier for customers to provide instantaneous feedback about the performance of remote employees and for firms to track that performance. Cheap surveillance cameras have made it easier to detect theft, sabotage, and other forms of bad conduct in the workplace. The spread of electronic payment mechanisms probably reduces opportunities to embezzle cash. These developments make it easier for employers to detect and deter shirking and other worker conduct that harms productivity and profitability. As a result, the labor rent premium falls according to efficiency-wage models founded on concerns about shirking and other hard-to-detect forms of worker misconduct.

Much anecdotal evidence points to the increased use of monitoring technologies to detect and deter shirking. My Google search of "employee monitoring technologies" on May 22, 2020, returned 139 million results. In the summary to its "Market Guide for Employee-Monitoring Products and Services," Gartner Research (2015) states that employee-monitoring tools "can protect sensitive information and generate positive ROI by increasing the productivity and efficiency of systems and employees. Security officers seek products and services in this market focusing on insider threat mitigation, regulatory compliance and employee productivity." While ubiquitous now, these technologies did not exist twenty years ago, or they existed only in more primitive and less capable forms. To my knowledge, however, economists have not studied their impact on rent premiums and wage structures. As I remarked above, the shirking-class of efficiency-wage models predicts that the spread of such technologies lowers rent premia.

It is also plausible that advances in monitoring technologies facilitate fissuring of the workplace by making it easier for firms to outsource noncore labor activities to other firms that specialize in those activities. This type of outsourcing relaxes internal pay equity constraints, leading to a loss of rents for low-pay workers. Fissuring also makes it easier for firms with market power and monopoly profits to de-link the compensation of noncore workers from firm-level profitability. The likely effect is to the reduce the rent premia of noncore workers and perhaps to raise them for core workers. The impact of advances in monitoring technologies on fissuring and labor rent premia is another topic that is ripe for research.

In closing, let me note that my remarks are not intended to deny a role for policy shifts and institutional forces in the decline of average labor rent premia. They may well play major roles, but the fall in unionization, for example, is not sufficient to make that case. As my foregoing remarks suggest, it seems likely that globalization and technological developments played important roles in driving the fall in unionization and average rent premia in recent decades. I also note that policies and institutions can affect how product market developments, globalization, automation, and advances in monitoring technologies have an impact on labor rent premia. So, similar exogenous developments may play out quite differently across countries with different policies and institutions.

REFERENCES FOR THE DAVIS COMMENT

Bollinger, Christopher R., Barry T. Hirsch, Charles M. Hokayem, and James P. Ziliak. 2019. "Trouble in the Tails? What We Know about Earnings Nonresponse 30 Years after Lillard, Smith and Welch." *Journal of Political Economy* 127, no. 5: 2143–85.

Bulow, Jeremy I., and Lawrence H. Summers. 1986. "A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment." *Journal of Labor Economics* 4, no. 3, pt. 1: 376–414.

Dutta, Sumair. 2012. "Field Service 2012: GPS and Fleet Management." Aberdeen Group White Paper. www.teletrac.com/teletrac.com/assets/aberdeen-research-pack.pdf.

Freeman, Richard B. 1980. "Unionism and the Dispersion of Wages." *Industrial and Labor Relations Review* 34, no. 1: 3–23.

Gartner Research. 2015. "Market Guide for Employee-Monitoring Products and Services." https://www.gartner.com/en/documents/2992419.

Gittleman, Maury, and Brooks Pierce. 2011. "Compensation for State and Local Government Workers." *Journal of Economic Perspectives* 26, no. 1: 217–42.

Hirsch, Barry, and David A. Macpherson. 2003. "Union Membership and Coverage Database from the Current Population Survey: Note." *Industrial and Labor Relations Review* 56, no. 2: 349–54.

Meyer, Bruce D., Walter K. C. Mok, and James X. Sullivan. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives* 29, no. 4: 199–226.

Philippon, Thomas, and Ariel Reshef. 2012. "Wages and Human Capital in the U.S. Finance Industry: 1909–2006." *Quarterly Journal of Economics* 127, no. 2: 1551–1609.

Shapiro, Carl, and Joseph Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review* 74, no. 3: 433–44.

Smith, Matthew, Owen Zidar, and Eric Zwick. 2020. "Top Wealth in America: New Estimates and Implications for Taxing the Rich." Working Paper. http://ericzwick.com/wealth/wealth.pdf.

COMMENT BY

CHRISTINA PATTERSON Anna Stansbury and Lawrence Summers provide a great and thought-provoking paper bringing together an array of results that paint a compelling case for the importance of the fall in labor rents in explaining several of the trends in the macroeconomy over the past several decades. The list of these important trends include a fall in the labor share, a rise in Tobin's q, a fall in measured markups, weak investment relative to profits, and a fall in steady-state unemployment. This paper posits that a fall in worker bargaining power is the major structural change that is responsible for all of these movements. The paper has two main parts—first, the authors measure the decline in the rents going to labor and demonstrate that the fall is large enough to explain many of these aggregate trends. Second, the authors present several cross-industry results that lend support for the predictive power of the fall in labor rents for all of these phenomena.

The paper has an abundance of evidence supporting the link between the fall in worker bargaining power and each of the macroeconomic phenomena mentioned above. I will focus on the relationship between the fall in labor rents and one of these trends—the fall in the labor share—to highlight some features that make the fall in labor rents a compelling explanation for this phenomenon.

From my reading of the literature, there are several features of the fall in the labor share that any explanation must accommodate. First, the fall in the labor share is largest for manufacturing, and it began several decades before the fall in other sectors (Gutiérrez and Piton forthcoming). Second, the fall in the labor share occurred mostly within sectors and is not primarily the artifact of a reallocation of economic activity from manufacturing to services (Elsby, Hobijn, and Şahin 2013). Third, some elements of the fall in the labor share are international in scope, although there is important variation across countries and the fall is particularly pronounced within the United States (Karabarbounis and Neiman 2014). Fourth, the fall in the labor share did not reflect a fall in the labor share at the median firm but rather is the result of a reallocation of economic activity within sectors to large firms with lower and potentially falling labor shares (Autor and others 2020).

A particularly compelling and exciting feature of the Stansbury and Summers explanation for the fall in the labor share is that it is consistent with all of these empirical findings. For example, we know that the decline in union membership was particularly pronounced in manufacturing, where the fall in the labor share was the largest. The authors show in the paper that the fall in labor rents occurred predominately within sector, consistent with a within-sector fall in the labor share. We also know that declines in union membership and the rise in domestic outsourcing are widespread in the developed world, suggesting perhaps that the fall of worker bargaining power is as well. And, as the authors note in section V, big firms generally have more rents and so they may disproportionately benefit from falling worker power, making the fall in worker bargaining power potentially consistent with the between-firm reallocation patterns that characterize the fall in the labor share.

The rest of this discussion will focus primarily on the authors' measurement of labor rents since this measurement of the decline in labor rents is at the core of the analysis and unpacking this measure will shed light on why labor rents have fallen and what it could mean for policy. The authors measure total rents going to labor as the sum of industry rents, union rents, and firm size rents. They get these estimates by jointly estimating the following regression relating the hourly wage (w_{it}) to individual i's characteristics and a set of industry, union, and firm size fixed effects:

$$\log(w_{it}) = \underbrace{X'\beta}_{\text{individual}} + \underbrace{\gamma_{ind} + \gamma_{union} + \gamma_{size}}_{\text{job}} + \epsilon_{it}. \tag{1}$$

The basic intuition is that rents are differences in worker earnings that are not explained by the demographics or occupation of the worker ($X'\beta$) but are explained by the characteristics of the job in which the individual works (γ_{ind}, γ_{union}, and γ_{size}). This strategy is quite reasonable and builds on a long line of literature using related estimates to infer the amount of rent-sharing in the economy (Katz and Summers 1989). This regression is also closely related to more recent literature estimating the importance of firms in explaining worker earnings. Firm fixed effects in these papers are derived from a very similar regression to the one that Stansbury and Summers run in their paper, but disaggregating the industry, union, and firm size fixed effects into a firm fixed effect:

$$\log w_{ijt} = \underbrace{X'\beta + \gamma_i}_{\text{individual}} + \underbrace{\gamma_j}_{\text{job}} + \epsilon_{it}, \tag{2}$$

where j is the firm and i is the individual. Under a similar logic to the analysis above, wherein a fall in the variance of the job-level fixed effects in equation (1) signals a fall in labor rents, a fall in the variance of firm fixed effects in equation (2) would also suggest a fall in worker rents. Song and others (2019) use data from the IRS to provide estimates for the variance of firm fixed effects over time in the United States. They document that the total variance of earnings rose from 1980 to 2013 but that the variance of firm fixed effects was relatively flat and fell only slightly, from 0.084 to 0.081, over this period. Rather, they find that the rise in the variance in worker earnings was largely explained by a rise in the variance of individual fixed effects and a rise in the covariance between worker fixed effects and firm fixed effects.

Initially, the findings of Song and others (2019) seem at odds with the steep fall in industry fixed effects and worker bargaining power in figures 3 and 5 in Stansbury and Summers's paper. What might reconcile a fall in the variance of industry fixed effects in this paper with a relatively stable variance of firm fixed effects? Of course, one answer to this question is that these are different data sets; Stansbury and Summers's estimates are derived from a monthly survey with a short rotating panel while the estimates in Song and others (2019) are from annual tax return data, and thus issues with measurement error, variable definitions, or sample selection could be driving the differences. Indeed, in recent work, Haltiwanger and Spletzer (2020) use administrative census data and find a rising variance of industry fixed effects, suggesting that a careful examination of the relationship between these data sets and estimates is warranted.

However, there is another more fundamental and, I would argue, more interesting explanation for the seemingly disparate patterns in the variance of firm fixed effects and industry fixed effects. Since all of the estimates here are log-additive, there is a simple relationship wherein the industry fixed effect (γ_{ind}) is simply the weighted sum of all the fixed effects (γ_j) of the firms within that industry (i.e., $\gamma_{ind} = \sum_j \frac{E_j}{E_{ind}} \gamma_j$). The industry fixed effect can fall either because the firm fixed effects fall or because there are more workers at the low fixed effects firms; in other words, the variance of industry fixed effects could be falling not because the fixed effects of the firms are falling (i.e., each firm is changing the rents that it gives to its workers) but rather because more workers are at the firms with the low fixed effects. This mechanism, wherein workers are being concentrated in low fixed effects firms, echoes the importance of reallocation in explaining the fall in the labor share, as I discussed above.

Aside from reallocation being a mechanism that reconciles the disparate estimates in the literature, there is also direct evidence that the fall in labor rents is at least partially the result of workers reallocating across firms. In particular, the evidence demonstrates that the reallocation of workers was nonrandom, meaning that in addition to a falling total amount of rents going to workers, there was also a meaningful reallocation of these rents *across* workers. One piece of evidence suggesting this reallocation of labor rents can be seen in the finding by Song and others (2019) that the covariance between worker effects and firm effects played an important role in explaining the rise in the variance of worker earnings, suggesting that high-ability workers are now more likely to be in the higher-rent firms. Additionally, recent work by Kline and others (2019) shows that, within the firm, rent-sharing is concentrated in the top half of the earnings distribution. Evidence in this paper and from Bloom and others (2018) shows that the large firm wage premium has fallen for low-education workers but has stayed the same or possibly even risen for higher-education workers. And last, we know that there has been a growth in domestic outsourcing, which is concentrated among the lower-skill occupations and results in significant wage losses for those who are displaced (Goldschmidt and Schmieder 2017). Taken together, this evidence strongly supports the authors' finding in section II.A that labor rents have fallen substantially farther for non-college-educated workers than for college-educated workers. Moreover, it suggests to me that a potential mechanism behind the fall in labor rents that they measure in this paper is that low-skill workers are concentrating in low-rent firms. This force leads to both a fall in the total rents going to workers and a reallocation of those rents across workers.

Understanding the mechanism behind the decline in labor rents is a very important next step in understanding the importance of the decline in worker rents for several reasons. First, this is a paper about providing a unified explanation for several macro phenomena, and this reallocation mechanism extends the analysis to directly link the fall of labor rents to the rise in income inequality, another salient feature of the economy over the past decades. The current analysis highlights inequality stemming from differences in capital and labor income, but we know that a large component of the rise in income inequality comes from changes in the distribution of labor income (Smith and others 2019). A fall in labor rents that is driven by a reallocation of low-skill workers to low-rent firms intrinsically links the fall in total labor rents to the reallocation of labor rents across workers.

Second, recognizing that the fall in labor rents is concentrated among some groups of workers strengthens the connection that the authors make between the overall reallocation of rents and several labor market trends, many of which we know are also concentrated among the low-skill workers. This is evidenced by the authors' finding in section IV.C that the natural rate of unemployment has fallen farther for workers without a college degree than for those with a college degree, which echoes the patterns in Crump and others (2019). Similar cross-demographic patterns are likely also true for the decline in labor force participation or the growth in real wages.

Third, understanding the mechanism that produced the estimated fall in labor rents is important because it suggests some candidates for *why* labor rents have fallen and thus what policies are best suited to addressing this phenomenon. A fall in labor rents that is driven by a fall in the share of workers at high-rent firms or by a concentration of rents within the firms that employ high-skill workers suggests that some explanations are more likely than others. The authors posit that the decline in worker power could be the result of institutional changes, the result of changes in priorities within firms, or the result of a change in economic conditions that make it harder for workers to bargain. However, a fall in labor rents driven by reallocation suggests that a simple fall in the bargaining power of workers within the firm may be incomplete. Perhaps it is not that the workers at Apple, for example, are not sharing in rents but rather that at Apple there are few workers relative to their share of aggregate value-added and most of the workers are high-skill. Moreover, the janitors at Apple may now technically work for a different firm, and that firm may not have any rents over which to bargain. In this case, policies that give the low-wage workers more bargaining power may not result in the workers getting more rents. And finally, if the fall in worker rents really is driven by the reallocation of workers or rents across firms, this could also suggest that rather than the fall in rents being the structural force that *caused* these many phenomena, it could itself be the *consequence* of some technological process that changed the nature of production and reallocated workers and economic activity across firms. Exploring this pattern is an important area for future research.

REFERENCES FOR THE PATTERSON COMMENT

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics* 135, no. 2: 645–709.

Bloom, Nicholas, Fatih Guvenen, Benjamin S. Smith, Jae Song, and Till von Wachter. 2018. "The Disappearing Large-Firm Wage Premium." *American Economic Association Papers and Proceedings* 108:317–22.

Crump, Richard, Stefano Eusepi, Marc Giannoni, and Ayşegül Şahin. 2019. "A Unified Approach to Measuring *u**." *Brookings Papers on Economic Activity*, Spring, 143–238.

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin. 2013. "The Decline of the US Labor Share." *Brookings Papers on Economic Activity*, Fall, 1–63.

Goldschmidt, Deborah, and Johannes F. Schmieder. 2017. "The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure." *Quarterly Journal of Economics* 132, no. 3: 1165–1217.

Gutiérrez, Germán, and Sophie Piton. Forthcoming. "Revisiting the Global Decline of the (Non-housing) Labor Share." *American Economic Review: Insights*.

Haltiwanger, John C., and James R. Spletzer. 2020. "Between Firm Changes in Earnings Inequality: The Dominant Role of Industry Effects." Working Paper 26786. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26786.

Karabarbounis, Loukas, and Brent Neiman. 2014. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129, no. 1: 61–103.

Katz, Lawrence F., and Lawrence H. Summers. 1989. "Industry Rents: Evidence and Implications." *Brookings Papers on Economic Activity: Microeconomics*, 209–90.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar. 2019. "Who Profits from Patents? Rent-Sharing at Innovative Firms." *Quarterly Journal of Economics* 134, no. 3: 1343–1404.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. "Capitalists in the Twenty-First Century." *Quarterly Journal of Economics* 134, no. 4: 1675–1745.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2019. "Firming Up Inequality." *Quarterly Journal of Economics* 134, no. 1: 1–50.

GENERAL DISCUSSION Valerie Ramey started the discussion off by praising the paper and relating it to her work with Borjas which studied the increase in the college wage premium relative to the high school wage premium.[1] Their argument was that because US companies had more market power worldwide in the '50s and '60s, labor unions were more powerful since there were more rents that could be shared. As a result, workers with high school degrees earned relatively higher wages. Furthermore, she explained that they had linked this explanation to the durable goods industries and to the interindustry wage differentials explored by Krueger and Summers.[2]

Furthermore, Ramey pointed out that when she and Borjas published their paper there was far less competition from workers in other countries. In contrast, there is far more competition from workers from other countries today, and this has led to a decline in worker bargaining power. She concluded by emphasizing the importance of Stansbury and Summers's documentation of this more recent development.

John Haltiwanger noted that a key finding in the paper is a declining interindustry earnings differential. The analysis in the paper uses the Current Population Survey (CPS). However, his recent work with James Spletzer shows that this result is not robust to matched employer-employee administrative data.[3] In fact, they attribute most of the rising overall

1. George J. Borjas and Valerie A. Ramey, "Foreign Competition, Market Power, and Wage Inequality," *Quarterly Journal of Economics* 110, no. 4 (1995): 1075–1110.

2. Alan B. Krueger and Lawrence H. Summers, "Efficiency Wages and the Inter-Industry Wage Structure," *Econometrica* 56, no. 2 (1988): 259–93.

3. John C. Haltiwanger and James R. Spletzer, "Between Firm Changes in Earnings Inequality: The Dominant Role of Industry Effects," Working Paper 26786 (Cambridge, Mass.: National Bureau of Economic Research, 2020).

variance of earnings across individuals to sharply rising interindustry wage differentials. He then asked the authors how they had addressed the limitations of industry codes in household surveys and whether using survey data introduced measurement problems in their study.

Bob Hall discussed some of the implicit policy recommendations he believed could be found in the paper, such as encouraging unionization; banning activist shareholders, layoffs, and outsourcing; raising the minimum wage; and giving labor a voice in management. He agreed with Patterson's remarks during the discussion that while this package of reforms addresses some real problems, it doesn't recognize a deeper problem in the labor market: the appalling condition of workers with low education. Policies such as the minimum wage create poverty traps by excluding low-skill individuals from employment.

Henry Aaron noted the conspicuous absence of any explicit reference to unionization in their presentation. In his view, the most striking fact from recent history is not so much the decline of previously unionized industries but the failure of unionization to proceed at all in emerging industries. Therefore, he asked the authors and discussants for their opinion on the extent to which hostility toward unionization is a political factor in addition to the economic trends they had discussed already.

Katharine Abraham commended the paper and posed another measurement question. She observed that the authors' measure of labor rents was constructed from information on the effects of being a union member, working in a particular industry, and working for a firm of a particular size. Another relevant feature of the labor market is occupational licensing, something that has grown in importance in recent decades. Abraham pointed to research suggesting that occupational licensing may lead to labor rents for workers in those occupations.[4] She asked the authors how accounting for this would impact their conclusions.

Daron Acemoglu found the paper intriguing and remarked that the issue of worker power is very interesting. However, he expressed concern about omitted variable bias stemming from the association of worker power with technological changes, offshoring, outsourcing, and automation. In particular, each of these factors directly impacts the labor share and wage

4. Morris M. Kleiner and Alan B. Krueger, "Analyzing the Extent and Influence of Occupational Licensing on the Labor Market," *Journal of Labor Economics* 31, no. 2 (2013): S173–S202; Maury Gittleman, Mark A. Klee, and Morris M. Kleiner, "Analyzing the Labor Market Outcomes of Occupational Licensing," *Industrial Relations* 57, no. 1 (2018): 57–100; Marc T. Law and Mindy S. Marks, "The Labor-Market Effects of Occupational Licensing Laws in Nursing," *Industrial Relations* 56, no. 4 (2017): 640–61; Robert J. Thornton and Edward J. Timmons, "Licensing One of the World's Oldest Professions: Massage," *Journal of Law and Economics* 56, no. 2 (2013): 371–88.

inequality—especially for low-wage workers. Given that their measure of imputed rents is a realized measure, any of the factors impacting the labor share he mentioned would simultaneously influence the variable that their paper tried to explain and their measure of imputed rents. Therefore, he wondered whether their approach exaggerated the importance of worker power and suggested that they try to put bounds on this issue.

Larry Katz said that while he found their paper to be very insightful, he noted that despite showing a strong difference for the firm size wage premium by education, the authors didn't make this distinction for their analysis at the industry level. He suggested that they decompose their results on labor rents by educational group, since he believed this could lead them to more granular insights on the decline in the labor share. Furthermore, he remarked that the authors' findings could help explain the decline in labor force participation, since it may be that the decline in rents for less-educated workers has decreased unemployment by leading workers to drop out of the labor market.

Greg Mankiw wondered whether declining worker power is a problem that needs to be addressed. Mankiw stated that worker power is synonymous with monopoly on the part of labor and alluded to the various problems economists have linked to monopoly power. He suggested that if it is true that worker power has declined, perhaps instead of reversing this trend we should applaud it as it might signal that labor markets have become more competitive. If the main concern with this trend is related to inequality, he suggested that it might be better to implement redistribution through tax policy rather than by increasing worker power.

Hall replied to Mankiw's comment by emphasizing that he didn't have a policy agenda that responds to these problems and that he thought the paper was not very specific on this point either. In his view, a lot of the discussion alluded to a national crisis related to the conditions of workers with low education. He found it surprising that this paper implicitly suggested that embracing old-fashioned continental European policies is the answer. However, he stated that making labor markets completely competitive in every regard has not worked very well either and has done a particularly terrible job for low-income people. He wished to make it very clear that in his view, Thatcherism is not the answer to these problems.

Larry Summers thanked the participants and stated that he couldn't remember a more thoughtful set of comments on a paper he had presented at Brookings. One phrase that was very present in his mind which was not mentioned in the paper was "ruthless economy," which refers to the idea that workers were being driven increasingly to marginal productivity because of external competition and declining unionization. He sympathized with the comments that went beyond the scope of their paper on the implications of declining worker power for workers with different education backgrounds and workers in different industries. However, he also agreed that a better understanding of the causal mechanisms behind this would be very helpful. In addition, he agreed that they definitely need to look into the measurement concerns that were brought up.

He acknowledged that the changes in the legal framework that complicated union organizing were an important part of understanding what has happened regardless of one's view on whether this has been good or bad. Moreover, he admitted that they had been intentionally vague about drawing policy implications from their paper because they wished to be highly cautious of embracing a continental European or Thatcherite agenda. He concluded by saying that while their paper didn't fully resolve our understanding of the plight of less-educated American workers, he believed their contribution was still valuable as it put movements in various aspects of labor power high up in the agenda for research on major economic trends.

Anna Stansbury added that the measurement issues raised were very helpful. With regards to Haltiwanger's comment, her understanding was that the rise in wage differentials between industries was related to increased sorting of high-wage workers into higher wage industries and firms, and therefore the industry and firm fixed effect hadn't increased over time. Nevertheless, she was eager to follow up with the commenters on how they could reconcile their estimates with these measurement problems as interindustry wage differentials are a central part of their analysis.

While the main focus of their paper was not to study educational divides, she replied to Katz's suggestion by saying that they started looking into this more recently. She also noted that their online appendix does include a short section showing how these trends have hit low-education workers especially hard. Moreover, they found that the decline in unionization in the private sector has been concentrated on workers with no college degree and some evidence that the fall in the dispersion of industry rents was greater for lower-education workers.

Last, she agreed that the reallocation between firms is a very important part of their story but suggested that it could also be the case that the decline in worker power could be partly causing some of the reallocation toward high markup firms. She reasoned that a decline in worker power would disproportionately reduce labor costs for high-rent firms—since there were more rents being shared ex ante—enabling them to expand relatively more quickly.

KATHARINE G. ABRAHAM
University of Maryland

JOHN C. HALTIWANGER
University of Maryland

LEA E. RENDELL
University of Maryland

How Tight Is the US Labor Market?

ABSTRACT We construct a generalized measure of labor market tightness based on the ratio of vacancies to effective searchers. Our generalized measure exhibits substantially less volatility than the standard measure defined as the ratio of vacancies to unemployment. Effective searchers include not only the unemployed but also those who are out of the labor force and the employed. These groups account for a substantial share of hires and their presence mutes the effects of the pronounced countercyclical movements in unemployment. The effective searcher measure also distinguishes different groups among the unemployed. During protracted contractions, the distribution of unemployment shifts toward the long-term unemployed, a group with lower relative search intensities, contributing to the smaller proportional increase in effective searchers as compared to the simple unemployment count. The Beveridge curve constructed using effective searchers is much more stable than the standard Beveridge curve. Further, the matching function for hires based on our generalized measure outperforms the matching function based on the ratio of vacancies to unemployment. Our approach thus reduces the unexplained residual variation required in the matching function to be consistent with real world data.

Conflict of Interest Disclosure: Katharine G. Abraham is professor of economics and survey methodology at the University of Maryland; John C. Haltiwanger is the Dudley and Louisa Dillard Professor of Economics and Distinguished University Professor of Economics at the University of Maryland; Lea E. Rendell is a graduate student in economics at the University of Maryland. Beyond these affiliations, the authors did not receive financial support from any firm or person for this paper or from any firm or person with a financial or political interest in this paper. They are currently not officers, directors, or board members of any organization with an interest in this paper. No outside party had the right to review this paper before circulation. The views expressed in this paper are those of the authors and do not necessarily reflect those of the University of Maryland.

For a simple summary of labor market conditions, observers and analysts long have turned to the unemployment rate. Unemployment exhibits clearly cyclical behavior, rising during downturns and falling during recoveries. By this metric, the labor market was tighter at the end of 2019 than it had been for half a century. Search and matching models of the labor market (Diamond 1982; Blanchard and Diamond 1989, 1992; Mortensen and Pissarides 1994; Pissarides 2000) imply that unemployment (or more generally job searchers) should be considered together with job openings in assessing labor market tightness. In these models, a higher ratio of job openings to unemployment makes it harder to recruit workers and easier to find jobs, thus indicating that the labor market is tighter. Official US statistics on job openings are available only since 2000, but the ratio of vacancies to unemployment at the end of 2019 was substantially higher than at any point since they began to be collected.

While both the unemployment rate and the ratio of vacancies to unemployment imply that the labor market was extraordinarily tight as of the end of 2019, there is reason to question whether these standard measures in fact capture the true degree of labor market tightness. This paper will focus on the measure at the heart of conventional search and matching models—the ratio of vacancies (V) to unemployment (U). In the search and matching framework, both the job-filling rate and the job-finding rate should vary systematically with V/U. In fact, however, these rates often deviate substantially from the rates implied by calibrated matching functions with V/U as the driving variable. This was especially true during and following the Great Recession, a finding often explained by an appeal to unobserved fluctuations in matching efficiency (Elsby, Michaels, and Ratner 2015). One of this paper's central goals is to explore whether and to what extent the use of broader measures of effective vacancies and effective searchers in the matching function can improve its fit.

One reason V/U might perform poorly in tracking job-filling and job-finding rates is that the number of unemployed people may be a poor proxy for the availability of workers to fill vacant jobs. For one thing, different groups among the unemployed may be more or less attached to the labor market and more or less likely to move into employment. In a seminal paper published in *Brookings Papers on Economic Activity*, Perry (1970) noted that women's rising labor force participation and the entry of the baby boom generation into the labor force could have raised measured unemployment independently of underlying labor market conditions. In recent years, researchers have argued that the higher-than-usual share of long-term unemployment among the unemployed following the Great

Recession implied effective unemployment lower than suggested by the unemployment rate (Krueger, Cramer, and Cho 2014). Further, as noted by Blanchard and Diamond (1989) and emphasized in a number of recent empirical studies, most new hires originate from out of the labor force or from another job (job-to-job flows) rather than from unemployment. The standard model ignores these job seekers altogether. Another potentially important factor is that the search intensity of potential workers in any given category may vary over time (Krueger and Mueller 2010, 2011; Davis 2011; Hall and Schulhofer-Wohl 2018).

Simply looking at job vacancies may be problematic as well. Existing evidence implies that the intensity with which firms recruit to fill their vacancies varies over time. Along with doing a better job of measuring the number of effective searchers, accounting for this variation also could help to account for observed variation in both job-filling and job-finding rates (Davis, Faberman, and Haltiwanger 2013; Gavazza, Mongey, and Violante 2018; Mongey and Violante 2020).

Building on the search and matching literature, including several prior studies that have made use of augmented measures of either effective searchers or effective vacancies, we propose a generalized measure of labor market tightness that addresses some important limitations of the standard measures. We begin our analysis by broadening the concept of effective job searchers to account not only for job candidates drawn from among the unemployed but also from among those currently out of the labor force or already working, as well as for the heterogeneity within each of these three groups of searchers. In defining the groups used for our analysis, we disaggregate the population relatively finely by labor market status in order to capture as much of the heterogeneity in job search behavior as possible. We construct a measure of effective searchers by taking a weighted sum across twenty-two different groups within the population age sixteen and older, with the weights based on the relative base period job-finding rates for each of the different groups as a proxy for relative job search intensities.

This generalized measure of effective searchers, expressed as a share of the population, exhibits much less volatility than the unemployment rate constructed on the same basis. One reason is that searchers who are out of the labor force or employed mute the effects of the pronounced countercyclical movements in unemployment on the broader generalized measure. Second, composition effects matter. In particular, during a deep and extended contraction such as the Great Recession, the composition of unemployment shifts toward the long-term unemployed, a group with lower relative job-finding rates. This further dampens the increase in the number of effective searchers in the broader measure during economic downturns.

One of the puzzles of recent labor market history has been the pronounced and persistent outward shift in the Beveridge curve relating job vacancies and unemployment following the Great Recession. We find that over the period from 1994 through 2019 the Beveridge curve constructed using vacancies and effective searchers is much more stable than the curve constructed using vacancies and unemployment. Further, using our generalized measure of effective searchers rather than unemployment in the measure of labor market tightness reduces the unexplained residual variation in the job-filling rate (hires per vacancy) and in the job-finding rate of unemployed workers (hires from unemployment per unemployed person). Over the 1994–2019 period, the residual unexplained variation in the job-filling rate calibrated using our generalized measure of effective searchers is only about half as large as that based on the standard measure. For the job-finding rate among the unemployed, the residual unexplained variation of the calibrated series using our generalized effective searchers measure is about a third lower than that based on the standard measure.

Our baseline analysis is conservative in that it neglects some of the important factors that a generalized measure of labor market tightness ideally would take into account. First, the relative intensity of search on the part of job seekers in any given group may change over time. Second, the intensity with which employers recruit to fill their vacant jobs also may change over time. We extend our baseline analysis with an exploratory investigation of time-varying relative search and recruiting intensities. Given the measurement challenges associated with this exercise, we view the results as suggestive rather than conclusive. Even so, our efforts further reduce the unexplained residual variation in job-filling rates. The evidence regarding the job-finding rates of the unemployed is mixed. Our accounting for time-varying recruiting intensity reduces the unexplained residual variation in job-finding rates, but our accounting for time-varying search intensity actually increases it. Better measures of both search intensity and recruiting intensity could well produce better results.

I. A Broader Perspective on Labor Market Tightness

A matching function that relates hires to unemployment and vacancies is the cornerstone of modern macro models of the labor market. Especially in recent years, however, it has been necessary to posit substantial fluctuations in matching efficiency to account for observed fluctuations in job finding and job filling. Fluctuations in matching efficiency similarly have been offered as an explanation for shifts in the Beveridge curve that describes the inverse relationship between vacancies and unemployment (Elsby, Michaels, and Ratner 2015).

An alternative approach that we build on here is to consider whether using more suitable arguments in the matching function can eliminate the apparent instability of the standard model. Rather than equating effective searchers with the unemployed, this broader perspective on labor market tightness recognizes that the people available to fill open jobs also include those out of the labor force or in another job. Constructing an aggregate measure of effective searchers also requires a way to measure the search intensity of those in the effective searcher pool. Relatedly, a broader perspective should recognize that labor market tightness depends not only on the number of posted job openings but also on how hard employers are trying to fill those jobs.

I.A. The Pool of Effective Searchers

A lengthy literature has examined how changes in the composition of the unemployed may affect the interpretation of the official unemployment rate. One strand of the literature, launched by the seminal work of Perry (1970) and further developed by Shimer (2001), Aaronson and others (2015), and Barnichon and Mesters (2018), among others, focuses on the demographic composition of the unemployed. Another strand focuses on the relative numbers of long-term and short-term unemployed (Kaitz 1970; Krueger, Cramer, and Cho 2014). Whether because of lower search intensity, loss of human capital, or employer unwillingness to hire them (Abraham and others 2019), the long-term unemployed have lower job-finding rates and may contribute proportionately less than the short-term unemployed to the pool of effective searchers. How a person entered unemployment also may be important. As an example, the job-finding pattern among those laid off from a job differs considerably from the pattern for other groups among the unemployed (Katz 1986; Katz and Meyer 1990; Fujita and Moscarini 2017).

A comprehensive measure of effective job searchers also needs to account for people who are outside of the labor force. Those in this group have a much lower job-finding rate than the average unemployed person, but there are many more of them. In a typical month, the number of people who enter employment directly from out of the labor force is much larger than the number entering directly from unemployment

(Hornstein, Kudlyak, and Lange 2014). People who are out of the labor force but say they want a job are much more likely to enter employment than the rest of the out-of-the-labor-force population (Blanchard and Diamond 1989; Jones and Riddell 1999; Hornstein, Kudlyak, and Lange 2014; Kudlyak 2017; Hall and Schulhofer-Wohl 2018).

The employed are a final group of searchers (Sedlacek 2016; Hall and Schulhofer-Wohl 2018). In the canonical search-and-matching model, vacancies include the job openings created by departing employees. Symmetrically, the measurement of effective searchers should take into account on-the-job searchers who may fill those jobs. Available survey data suggest that on-the-job search is prevalent (Black 1980; Blau and Robins 1990; Faberman and others 2017) and administrative data show that a large share of hires are people moving from one job to another (Haltiwanger, Hyatt, and McEntarfer 2018; Haltiwanger and others 2018).

I.B. Job Search Intensity

In addition to properly identifying the effective searcher population, a full accounting of effective search activity also needs to incorporate search intensity. Approaches used in the literature to do this include directly measuring search activities, making use of information on the gap between individuals' desired and actual hours, and the approach that we adopt—inferring relative search intensity from relative job-finding rates.

In an early effort to measure job search intensity directly, Shimer (2004) uses information from the Current Population Survey (CPS) on the number of different search methods reported by the unemployed as a search intensity proxy. Several studies, including DeLoach and Kurt (2013), Gomme and Lkhagvasuren (2015), and Mukoyama, Patterson, and Şahin (2018), measure search intensity among the unemployed using data from the American Time Use Survey (ATUS) on time devoted to job search. ATUS data are available only beginning in 2003, but Mukoyama, Patterson, and Şahin (2018) combine them with CPS data on job search methods to construct a longer search intensity series. They model the relationship of search time to the search methods reported by the unemployed, then use that estimated relationship to construct a longer search intensity series for the unemployed. Studies using ATUS data reach conflicting conclusions about whether search intensity among the unemployed is procyclical or countercyclical. Ahn and Shao (2017) use ATUS data to study the cyclicality of job search among the employed. Because the ATUS does not ask what respondents are doing while they are at work, ATUS measures of job search among the employed seem especially likely to miss at least some job search activity.

Faberman and others (2020) use data on the gap between desired and actual hours to proxy for job search intensity. They show that, in 2013–2015 data from the Survey of Consumer Expectations (SCE), this gap is correlated with a direct measure of search intensity. They use the SCE data to calculate the average difference between desired and actual hours for each of thirty-nine groups defined based on labor force status and demographic characteristics. Treating the gaps for defined groups as constant over time, the authors use CPS data to construct a time series of aggregate slack, which they define as the total gap between desired and actual hours divided by total desired hours.

Finally, job-finding rates have been used to proxy for search intensity. The simplest version of this approach uses group-specific job-finding rates in a base period to weight the number of people in each group to produce an aggregate measure of effective searchers. A notable example of this approach is the Federal Reserve Bank of Richmond's Hornstein-Kudlyak-Lange Non-Employment Index (NEI), which uses long-run average job-finding rates to aggregate its nine groups of effective searchers among the unemployed and those out of the labor force (Hornstein, Kudlyak, and Lange 2014; Kudlyak 2017).

Several studies, including Veracierto (2011), Hornstein and Kudlyak (2016), and Sedlacek (2016), have sought to infer the variation in within-group search intensities from changes in relative job-finding rates. Among the studies adopting this general approach, Hall and Schulhofer-Wohl (2018) offer the most comprehensive characterization of the job searcher pool, considering sixteen groups of job seekers—thirteen groups among the unemployed and two among those out of the labor force plus the employed.

The basic strategy in all of these studies is, in effect, to infer what is happening to group-specific search intensities based on how having more or fewer people in any given group affects the number of matches. If adding people to a group makes a larger than expected contribution to the number of matches realized when the labor market is tight, for example, procyclicality in search intensity is a plausible explanation.[1] A limitation

1. Alternatively, the cross-group differences that are the basis for the suggested inference about job search intensity could be attributable to differences in the pattern of the shocks experienced by different groups of searchers. This is a less parsimonious explanation, and it is not entirely apparent what the source of such shocks might be, though it cannot be ruled out.

is that cyclical variation in search intensity that is common across groups cannot be distinguished empirically from the elasticity of matching with respect to the (properly measured) ratio of vacancies to searchers in the standard matching function or common changes in matching efficiency. A modeler can hope to quantify changes in aggregate search intensity that result from changes in the relative sizes of groups with relatively procyclical or relatively countercyclical job-finding rates. The potentially more important changes in search intensity that are common across groups, however, cannot be quantified using this approach.

I.C. Time-Varying Employer Recruiting Intensity

A final factor missing from the standard search-and-matching model is employer recruiting intensity. Empirical implementations of the standard model use data on the number of job openings. The intensity with which employers recruit to fill their openings can vary considerably, however, depending both on the company's own circumstances and on aggregate labor market conditions.

The most literal interpretation of recruiting intensity is the time and effort devoted to advertising the firm's job openings, processing applications, and so on, but other aspects of firms' recruiting behavior may be even more important. As an example, in a tighter labor market, employers may choose to consider job candidates with criminal records (Casselman 2018; Smialek 2019), lower the levels of education and experience they require, offer better working conditions, or raise wages. All of these can be viewed as changes in recruiting intensity, in the sense that employers are trying harder to fill their vacant jobs.

Modestino, Shoag, and Balance (forthcoming) show that, controlling for occupation, the shares of online job advertisements stating a requirement for a college degree or four or more years of experience rose during the Great Recession. The changes were larger in states and occupations that experienced a larger increase in the supply of workers. Using establishment-level Job Openings and Labor Turnover Survey (JOLTS) data, Davis, Faberman, and Haltiwanger (2013) show that, holding aggregate conditions constant, employers with a larger gross hiring rate fill their jobs at a faster pace. They interpret this finding through the lens of recruiting intensity—that is, they infer that recruiting intensity is positively associated with the gross hiring rate. Their findings suggest that the decline in hiring rates during the Great Recession should have led to a corresponding decline in recruiting intensity. Cross-sectional heterogeneity in recruiting intensity also may imply changes over time in overall recruiting intensity attributable to composition effects (Davis, Faberman, and Haltiwanger 2012b). Building on the results of Davis, Faberman, and Haltiwanger (2013), Gavazza, Mongey, and Violante (2018) develop an equilibrium model of the recruiting intensity response to negative shocks. Their model incorporates both the gross hiring rate channel and a second channel in which, in a weaker labor market, firms of all types exert less effort to fill their jobs. Based on a calibration exercise, they conclude that the latter effect is more important than the former. Mongey and Violante (2020) argue similarly that the effect of labor market slackness on firms' chosen recruiting intensity is a key driver of the residual variation in match rates during the Great Recession.

II. An Organizing Framework

In the standard specification of the canonical search-and-matching model (Diamond 1982; Blanchard and Diamond 1989, 1992; Mortensen and Pissarides 1994; Pissarides 2000), employers create job openings they would like to fill (V), and unemployed individuals (U) search among these job openings for employment. The process of matching unemployed workers to vacant jobs is represented by a production function, often assumed to be Cobb-Douglas in form, with vacancies and unemployment as the inputs and matches (hires) as the output:

$$H_t = m(V_t, U_t) = \mu_t V_t^{1-\alpha} U_t^{\alpha}, \tag{1}$$

where H is hires, V is the number of job openings, U is the number of unemployed people, t is the time period, μ_t is a time-varying match efficiency parameter, and $\alpha(1 - \alpha)$ is the elasticity of the matching function with respect to unemployment (vacancies). In this framework, labor market tightness (θ_t) is expressed as:

$$\theta_t = \frac{V_t}{U_t}. \tag{2}$$

This relationship may be viewed through the lens of the job-finding rate, expressed as hires per unemployed worker:

$$\frac{H_t}{U_t} = \mu_t \left(\frac{V_t}{U_t}\right)^{1-\alpha} = \mu_t (\theta_t)^{1-\alpha}. \tag{3}$$

All else the same, when the labor market is tighter (when θ_t is larger), an unemployed individual is more likely to find a job. An alternative but equivalent approach is to view this relationship through the lens of the job-filling rate, expressed as hires per vacant job:

$$(4) \qquad \frac{H_t}{V_t} = \mu_t \left(\frac{U_t}{V_t} \right)^{\alpha} = \mu_t \left(\frac{1}{\theta_t} \right)^{\alpha}.$$

All else the same, when the labor market is tighter, an employer is less likely to be able to recruit an unemployed person to fill a vacant job.

Since the matching function as written in equation (1) has constant returns to scale, it can be expressed as a relationship among the hiring rate h, vacancy rate v, and unemployment rate u:

$$(5) \qquad h_t = m(v_t, u_t) = \mu_t v_t^{1-\alpha} u_t^{\alpha},$$

where $h = H/\mathrm{E}$, $v = V/\mathrm{E}$, and $u = U/\mathrm{E}$ and E is employment. An additional constraint is that, in steady state, the number of separations (inflows to unemployment) must equal the number of hires (outflows from unemployment). This steady state relationship can be expressed:

$$(6) \qquad \delta_t = h_t = m(v_t, u_t) = \mu_t v_t^{1-\alpha} u_t^{\alpha},$$

where δ is the separation rate (in this case separations from employment into unemployment expressed as a fraction of employment) and the other terms are as previously defined.[2] The downward sloping relationship between the unemployment rate and the vacancy rate implied by equation (6) commonly is termed the Beveridge curve.

In this framework, shifts in either δ or μ will shift the position of the Beveridge curve. Improvement in the matching function (an increase in μ), for example, shifts the Beveridge curve inward (lowering unemployment for given vacancies), while deterioration in the matching function shifts the Beveridge curve outward (raising unemployment for given vacancies). Shifts in μ also will affect the job-filling rate and the job-finding rate.

2. Nothing fundamental changes if this expression is modified to allow for steady state growth at rate g in desired employment, in which case the left-hand side becomes $\delta + g$. Although the standard Beveridge curve specification expresses vacancies and unemployment relative to employment, as in equation (6), for comparison with the generalized Beveridge curve examined later, our empirical implementation works with vacancies and unemployment relative to population.

In the model as just sketched out, $\theta_t = \frac{V_t}{U_t}$, but unemployment and vacancies are imperfect proxies for the measures of effective searchers and effective vacancies that we believe should be the objects of interest. Empirically, the standard framework requires substantial variation in μ_t to account for observed shifts in the Beveridge curve, as well as to explain the considerable residual variation between actual job-filling and job-finding rates and those implied by the model. One reason for this may be that, by ignoring heterogeneity among the unemployed and job search among those who are out of the labor force or employed, the standard specification misrepresents the stock of potential job candidates. If effective job seekers of each of the different types moved together with unemployment over time, it would not be important to account for them separately (Broersma and van Ours 1999; Sedlacek 2016), but this cannot be the case as increases in any one group imply decreases in others. In addition, the standard tightness measure does not account either for temporal variation in search intensity or for temporal variation in employer recruiting intensity.

We can elaborate the simple model to account for these complexities. Building on the standard hiring function, we can write:

$$H_t = m\left(\sum_j \rho_t^{v_j} V_{jt}, \sum_i \rho_t^{s_i} S_{it}\right) = \mu_t \left(\sum_j \rho_t^{v_j} V_{jt}\right)^{1-\alpha} \left(\sum_i \rho_t^{s_i} S_{it}\right)^{\alpha}, \tag{7}$$

where V_j represents the number of job openings from firms of type j, S_i represents the number of job searchers of type i, $\rho_t^{v_j}$ represents the intensity of employer recruiting effort for firms of type j at time t, and $\rho_t^{S_i}$ represents the intensity of job search on the part of searchers of type i at time t.[3] Note that the elasticity of the matching function specified in equation (7) as well as the unexplained residual variation in the actual number of matches as compared to the number implied by the model are likely to differ from those associated with the standard model specified in equation (1).

In this expanded framework, labor market tightness can be written as:

$$\tilde{\theta}_t = \frac{\sum_j \rho_t^{v_j} V_{jt}}{\sum_i \rho_t^{s_i} S_{it}}. \tag{8}$$

3. This formulation builds on specifications of generalized matching functions in Davis (2011), Davis, Faberman, and Haltiwanger (2013), and Abraham (2015).

We refer to the numerator of this expression as effective vacancies (EV_t) and the denominator as effective searchers (ES_t). The steady state equilibrium of hires equal to separations is now:

$$(9) \qquad \delta_t = h_t = m\left(\sum_j \rho_t^{v_j} v_{jt}, \sum_i \rho_t^{S_i} s_{it}\right) = \mu_t \left(\sum_j \rho_t^{v_j} v_{jt}\right)^{1-\alpha} \left(\sum_i \rho_t^{S_i} s_{it}\right)^{\alpha},$$

where separations are now all separations from employment and variables are rates expressed as fractions of the population. Over the course of a business cycle, absent changes in matching efficiency or other factors that shift the position of the generalized Beveridge curve, effective searchers and effective vacancies will move inversely as implied by equation (9).[4]

In this generalized setting, the job-filling rate is given by:

$$(10) \qquad \frac{H_t}{V_t} = \mu_t \left(\frac{1}{\tilde{\theta}_t}\right)^{\alpha} \rho_t^{v}, \text{ where } \rho_t^{v} = \frac{\sum_j \rho_t^{v_j} V_{jt}}{V_t}.$$

For the generalized model, the ratio of hires to unemployment is no longer a job-finding rate since not all hires come from among the unemployed. To characterize the job-finding rate for any subgroup or collection of subgroups, we build on the transformation of the matching function used by Hall and Schulhofer-Wohl (2018). Following Hall and Schulhofer-Wohl (2018), job-finding rates differ across groups only because of differences in their effective search intensity, meaning that they are the same on a search intensity–adjusted basis. The common search intensity–adjusted job-finding rate varies with the tightness of the labor market:

$$(11) \qquad \frac{H_t}{ES_t} = \frac{H_{it}}{ES_{it}} = \frac{H_{it}}{\rho_t^{S_i} S_{it}} = \tilde{f}_t = \tilde{f}_{it} = A_t T_t^{\eta},$$

where $T_t = V_t/H_t$ is average vacancy duration and A_t are any common time effects on job-finding rates not captured by vacancy duration. Equation (11)

4. Something we have not considered explicitly is the possibility of mismatch between vacant jobs and effective job seekers. Although commonly cited by business leaders and policy officials as an important contributor to unemployment, especially during periods when the labor market is weak (Abraham 2015), available evidence suggests that mismatch plays at most a modest role in explaining aggregate unemployment fluctuations (Şahin and others 2014; Crump and others 2019). In our framework, we will think of mismatch as captured by μ.

is a transformation of (7). To see this, define $A_t = \mu_t^{1+\eta}\,(\rho_t^v)^\eta$. Then with appropriate substitution we have:

$$(12) \qquad H_t = \mu_t \left(\sum\nolimits_j \rho_t^{v_j} V_{jt}\right)^{\eta/(1+\eta)} \left(\sum\nolimits_i \rho_t^{S_i} S_{it}\right)^{1/(1+\eta)},$$

where $\alpha = 1/(1+\eta)$. Returning to the job-finding rate for group i, we can write:

$$(13) \qquad \frac{H_{it}}{S_{it}} = f_{it} = \rho_t^{S_i} A_t T_t^\eta,$$

Equation (13) can be used to quantify the implied job-finding rate for the unemployed. Using the subgroups among the unemployed defined in the general model, we have:

$$(14) \qquad \frac{H_{ut}}{U_t} = \frac{H_{ut}}{S_{ut}} = \mu_t^{1+\eta}(\rho_t^v)^\eta (V_t/H_t)^\eta \sum\nolimits_{i\in u} (\rho_t^{S_i}\, S_{it}/S_{ut}).$$

The standard model with a single group and ρ_t^v equal to 1.0 is just a special case of equation (14) given by:

$$(15) \qquad \frac{H_{ut}}{S_{ut}} = \mu_t^{1+\eta}(V_t/H_t)^\eta.$$

Equations (7)–(15) lay out the aspirational general model that measures both effective vacancies and effective searchers, allowing for cross-sectional and time series variation in search intensity for searchers and recruiting intensity for vacancies. Our primary goal is to evaluate the extent to which this generalization overcomes some of the known limitations of the standard framework. We report empirical analyses that compare and contrast calibrations of the standard and general models, emphasizing the residual unexplained variation in the matching function associated with each model manifest in the Beveridge curve, job-filling rate and job-finding rate.

In our empirical analysis, we focus first on a simpler version of equations (7)–(15) that allows for cross-sectional variation in relative search intensities across groups but not for variation in those relative search intensities over time. That is, we begin with a specification in which $\rho_t^{S_i} = \gamma_i$ for all t and $\rho_t^{v_j} = 1$ for all j and t. Our goal in this portion of the analysis is to quantify how much allowing for a broader group of effective searchers can improve the performance of the matching function. Then we explore specifications that permit relative search and recruiting intensities to vary over time.

III. Creating a Measure of Labor Market Tightness Based on Effective Searchers

Our measures of effective searchers build on several earlier papers, including the research underlying the Federal Reserve Bank of Richmond's Non-Employment Index (Hornstein, Kudlyak, and Lange 2014; Kudlyak 2017) and most especially the work of Hall and Schulhofer-Wohl (2018). Similar to these other papers, we use CPS micro data to track flows across labor market states and from job to job. Our analysis makes use of job-finding rates for each of twenty-two groups—thirteen groups among the unemployed (as in Hall and Schulhofer-Wohl 2018); three among those who are out of the labor force but say they want a job and four among others who are out of the labor force (as in the Richmond Fed NEI); and two among the employed.[5] Table 1 shows the full list of twenty-two groups. Following Hall and Schulhofer-Wohl (2018), we adjust each group's job-finding rate to hold demographics constant at the group's 2005–2007 values. In addition, following Fujita, Moscarini, and Postel-Vinay (2019), we make an adjustment for a change in the procedures used to collect CPS data that otherwise would lead to an understatement in job-changing rates among employed individuals. Details of the construction of the demographically adjusted job-finding rates are provided in online appendix B. As already described, we interpret the cross-group variation in (demographically adjusted) job-finding rates as variation in search intensity.

Table 1 shows, for 2006 and 2010, the population shares of each of our twenty-two groups, together with each group's demographically adjusted job-finding rate. Those who were recently temporarily laid off have the highest rate and those not in the labor force (retired) the lowest rate. Involuntary part-time workers have a job-finding rate that is twice that of other employed people, though still relatively low compared to other identified groups. Not surprisingly, average job-finding rates fell between

5. The featured Richmond Fed index distinguishes only two groups among the unemployed (short-term and long-term) and does not include employed searchers, though there is a second version that allows for search among those working part-time for economic reasons. In their work, Hall and Schulhofer-Wohl (2018) distinguish only two groups among those out of the labor force (want a job and do not want a job) and treat the employed as a single group.

Table 1. Estimated Relative Job-Finding Rates

	Share	JFR	Rel. JFR	Rel. JFR (raw)
2006				
Unemployed: Recently left job	0.16	39.46	0.71	0.61
Unemployed: Recently permanently laid off	0.21	32.80	0.59	0.52
Unemployed: Recently temporarily laid off	0.23	55.22	1.00	1.00
Unemployed: Temporary job recently ended	0.12	38.63	0.70	0.60
Unemployed: Recently newly entered	0.11	21.26	0.38	0.33
Unemployed: Recently reentered	0.34	29.89	0.54	0.46
Unemployed: Left job months ago	0.15	27.86	0.50	0.43
Unemployed: Permanently laid off months ago	0.36	21.19	0.38	0.33
Unemployed: Temporarily laid off months ago	0.16	44.28	0.80	0.69
Unemployed: Temporary job ended months ago	0.13	26.03	0.47	0.40
Unemployed: Newly entered months ago	0.12	14.75	0.27	0.23
Unemployed: Reentered months ago	0.45	23.44	0.42	0.37
Unemployed: Long-term unemployed	0.43	17.41	0.32	0.27
Want Job: Discouraged	0.15	14.74	0.27	0.23
Want Job: Looked last 12 months	0.43	14.24	0.26	0.22
Want Job: Other	1.24	15.26	0.28	0.24
Not in Labor Force: In school	4.34	9.41	0.17	0.15
Not in Labor Force: Retired	15.51	1.56	0.03	0.02
Not in Labor Force: Disabled	4.67	1.96	0.04	0.03
Not in Labor Force: Other	7.44	8.87	0.16	0.14
Employed: Involuntary part-time	1.79	5.12	0.09	0.08
Employed: Not involuntary part-time	61.44	2.22	0.04	0.03
2010				
Unemployed: Recently left job	0.09	27.81	0.54	0.48
Unemployed: Recently permanently laid off	0.29	23.12	0.45	0.38
Unemployed: Recently temporarily laid off	0.28	51.80	1.00	1.00
Unemployed: Temporary job recently ended	0.13	32.88	0.63	0.56
Unemployed: Recently newly entered	0.12	12.65	0.24	0.22
Unemployed: Recently reentered	0.27	21.30	0.41	0.37
Unemployed: Left job months ago	0.16	19.29	0.37	0.32
Unemployed: Permanently laid off months ago	0.90	14.41	0.28	0.24
Unemployed: Temporarily laid off months ago	0.26	36.15	0.70	0.60
Unemployed: Temporary job ended months ago	0.24	20.06	0.39	0.33
Unemployed: Newly entered months ago	0.24	9.41	0.18	0.16
Unemployed: Reentered months ago	0.57	16.45	0.32	0.28
Unemployed: Long-term unemployed	2.14	10.92	0.21	0.18
Want Job: Discouraged	0.47	11.33	0.22	0.19
Want Job: Looked last 12 months	0.52	9.76	0.19	0.17
Want Job: Other	1.27	12.30	0.24	0.21
Not in Labor Force: In school	5.07	6.28	0.12	0.11
Not in Labor Force: Retired	15.56	1.41	0.03	0.02
Not in Labor Force: Disabled	5.17	1.42	0.03	0.02
Not in Labor Force: Other	7.26	6.76	0.13	0.12
Employed: Involuntary part-time	3.73	3.63	0.07	0.06
Employed: Not involuntary part-time	55.27	1.77	0.03	0.03

Source: Authors' calculations using CPS.

Notes: Job-finding rates (JFR) estimated using CPS survey data linking households month to month. Relative job-finding rate (Rel. JFR) calculated by dividing all job-finding rates by job-finding rate for unemployed recently laid off. Recently unemployed groups refer to those unemployed 0–4 weeks. "Unemployed months ago" refers to those unemployed 5–26 weeks. "Long-term unemployed" refers to those unemployed 27 weeks or more.

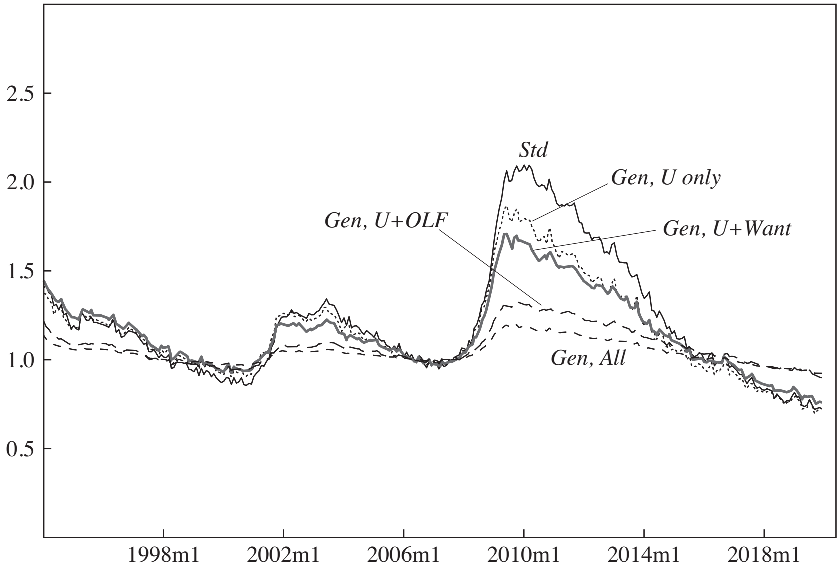
2006 and 2010 as the economy worsened following the onset of the Great Recession. Our primary interest, however, lies with the relative job-finding rates across the different groups. These are much more stable—the correlation in relative job-finding rates between 2006 and 2010 is 0.98. We construct our measure of effective searchers by weighting each of the twenty-two groups by its relative 2006 job-finding rate, shown in the third column of the top panel of table 1. All of the relative job-finding rates are defined with reference to the demographically adjusted 2006 job-finding rate of those who were recently temporarily laid off.

Table 1 also reports relative job-finding rates for 2006 and 2010 calculated directly from the CPS micro data not controlling for changing demographics, shown in the column denoted "Rel. JFR (raw)." Although our baseline analysis uses the demographically adjusted 2006 relative job-finding rate estimates, the results are very similar if we instead use the raw 2006 relative job-finding rates. Later in the paper, we present results based on demographically adjusted relative job-finding rates that are allowed to vary over time.

Figure 1 displays the standard measure of searchers (the simple unemployment count) along with several alternative measures that move toward our fully generalized effective searcher measure. The first of these considers only the unemployed as effective searchers but allows for changes in the composition of the unemployed across the thirteen groups we have specified. The second alternative measure incorporates the three groups among those out of the labor force who say they want a job, and the third adds the four groups among others who are out of the labor force. The fourth and final effective searcher measure is the fully generalized measure that adds the two groups of employed people. Each of the generalized measures weights the different groups it includes in accord with their relative 2006 job-finding rates.

The alternative series shown in figure 1 are highly correlated but distinctly different in their volatility. Allowing for heterogeneity among the unemployed yields a measure that is less cyclically volatile than the standard unemployment measure. Including, in turn, those who are out of the labor force but want a job, others who are out of the labor force, and finally the employed yields progressively less volatile measures. The first column of table 2 reports the standard deviations over the 1994–2019 period of the normalized series plotted in figure 1. Whereas unemployment as a share of the population, normalized to equal 1 in 2006, has a standard deviation of 0.34 over the 26-year period, the standard deviations of the alternative measures, also expressed as a share of the population and

Figure 1. Standard versus Generalized Measures of Searchers (Constant Relative Search Intensities)



Source: Authors' calculations using Current Population Survey (CPS).

Notes: All measures ratios to population age 16+ normalized to 1 in 2006. Std = unemployed; Gen, U only = generalized measure, unemployed only; Gen, U + Want = generalized measure, unemployed plus want a job; Gen, U + OLF = generalized measure, excludes employed; Gen, All = generalized measure, all twenty-two groups.

Table 2. Cyclical Volatility of Alternative Searcher Measures: Constant Relative Job Search Intensities

	Standard deviation, 1994–2019	*March 2001*	*June 2009*	*December 2019*
Standard	0.34	0.94	2.06	0.72
General, unemployed only	0.26	1.00	1.86	0.72
General, unemployed + want job	0.22	0.99	1.71	0.76
General, unemployed + OLF	0.10	0.97	1.32	0.90
General, all	0.06	0.99	1.19	0.92
U6	0.35	0.89	2.00	0.81
Richmond Fed NEI	0.11	0.96	1.34	0.89

Source: Authors' calculations using CPS and Richmond Fed NEI.

Notes: All measures normalized to 1 in 2006. Standard = unemployed; General, unemployed only = effective searchers with unemployed only; General, unemployed + want job = effective searchers with unemployed and want a job; General, unemployed + OLF = effective searchers excluding employed; General, all = all effective searchers; U6 = U6 measure of labor underutilization.

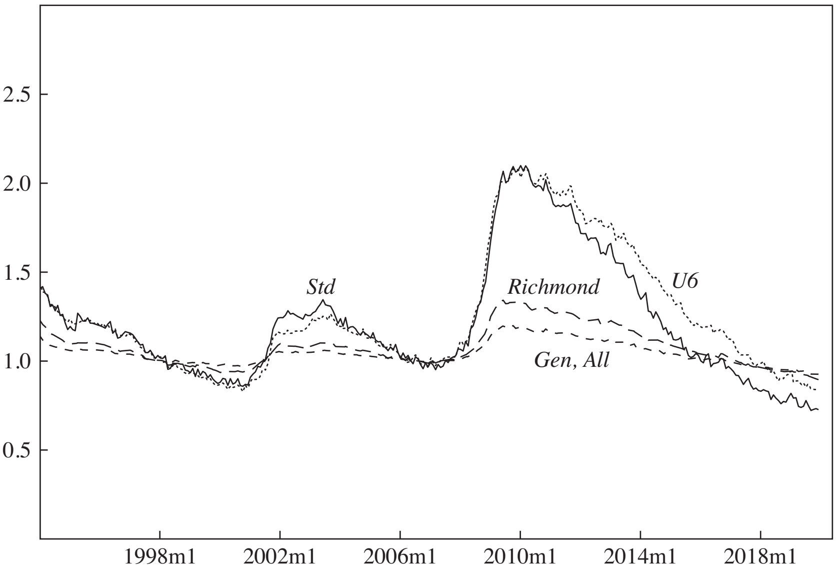
normalized to equal 1 in 2006, are progressively lower. The standard deviation of our fully generalized measure, calculated on a basis comparable to that of the unemployment measure, is just 0.06.

Another way to look at the alternative measures is to ask how the current level of effective searchers as a share of the population compares to the level at points of time in the past. The end of the prolonged expansion that lasted through most of the 1990s and into the early 2000s offers one interesting point of comparison. The standard measure of effective searchers (unemployment) relative to the population recorded in December 2019 was about 23 percent below its March 2001 value. In contrast, the December 2019 fully generalized measure of effective searchers relative to the population was only about 6 percent below its March 2001 value. Another interesting point of comparison is with the values for June 2009, the trough of the Great Recession. Whereas the standard measure of effective searchers fell by about 65 percent between June 2009 and December 2019, the fully generalized measure fell by just 23 percent.

One important reason for the muted cyclicality of the baseline generalized measure with fixed relative job search intensities as compared to the standard measure is that the generalized measure counts more people as effective searchers. The largest proportional fluctuations in group size over a typical cycle are the fluctuations among the unemployed. In the standard measure, any proportional increase in the number of unemployed people is de facto a proportional increase in the number of effective searchers. In the baseline version of the generalized measure, in contrast, the unemployed are only a fraction of all effective searchers, and increases in unemployment thus mechanically have a smaller proportional effect on the aggregate number of effective searchers. Composition also matters. When economic conditions are weak, the share of the unemployed who are long-term unemployed rises, and the long-term unemployed have lower job-finding rates than other unemployed people. In addition, in the baseline version of the generalized measure, even though the employed have lower relative search intensities than the unemployed, their search intensities are positive. This means that, during a downturn, reductions in the number of employed effective searchers partially offset increases in the number of effective searchers associated with rising unemployment.

Figure 2 compares our generalized measure of effective searchers with two alternative measures. The first is an index based on the Bureau of Labor Statistics U6 measure of slackness. The U6 measure counts the marginally attached and involuntary part-timers along with the unemployed but like

Figure 2. Standard, U6, Richmond Fed, and Generalized Effective Searcher Measures (Constant Relative Search Intensities)



Sources: Authors' calculations using CPS and Richmond Fed NEI.
Notes: All measures ratios to population age 16+ normalized to 1 in 2006. Std = unemployed; Gen, All = generalized measure, all twenty-two groups; U6 = U6 measure of labor underutilization; Richmond = Richmond Fed NEI.

the official unemployment rate weights all of them equally. The headline Richmond Fed NEI (Hornstein, Kudlyak, and Lange 2014; Kudlyak 2017) incorporates both the unemployed and those out of the labor force but distinguishes fewer groups among the unemployed than our generalized measure (two versus thirteen) and does not incorporate employed searchers. Similar to our measure, the different groups used in the Richmond Fed index are weighted based on persistent differences in their average relative job-finding rates. For consistency, we normalize both the U6 index and the Richmond Fed index by the population age sixteen and older.

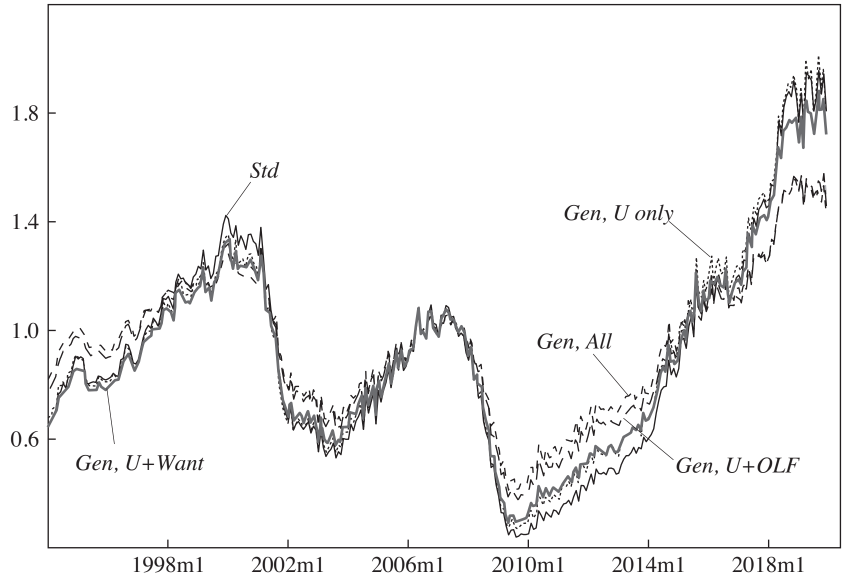
As can be seen in figure 2, the U6 index has about the same volatility as the standard measure. The cyclical variation in the Richmond Fed index is proportionally less than that of the standard measure but greater than that of our generalized index of effective searchers. Table 2 reports summary statistics for both the U6 and the Richmond Fed indexes. Consistent with the visual impression conveyed by figure 2, over the 1994–2019 period as a whole, the volatility of the U6 index is very similar to that of the unemployment index. The volatility of the Richmond Fed index lies between that of the standard measure and our generalized measure.

As already mentioned, in constructing our generalized measures of effective searchers, we have used estimated relative job-finding rates that hold demographics constant over time. We also have constructed similar measures using the simple average 2006 job-finding rates for the twenty-two groups. Comparing these measures, shown in online appendix figure A.1, to those in figure 1 suggests that controlling for demographics is relatively unimportant. As a further sensitivity analysis, we also have asked how much difference it makes that we have broken the unemployed into thirteen different groups, as opposed to distinguishing just between the short-term and the long-term unemployed. Measures constructed using the latter approach are shown in online appendix figure A.2. In addition, we have constructed an effective searcher measure that breaks the population into just five groups—short-term unemployed, long-term unemployed, out of the labor force and want a job, out of the labor force and do not want a job, and employed. The comparison between this measure and our fully generalized measure is shown in online appendix figure A.3. Over our time period, effective searcher measures constructed using these alternative approaches look broadly similar to measures constructed using our baseline approach. We note, however, that the similar behavior of these alternative searcher measures over this period does not necessarily imply they always will behave so similarly. There are marked differences in base period job-finding rates across the groups that are pooled together in the alternative measures, and, in a different period, distinguishing among them could make more of a difference. Later in the paper, we consider the performance of each of the alternative job searcher measures in explaining job-filling and job-finding rates.

We now are ready to compare labor tightness measured using effective searchers as opposed to unemployment. The numerator for all of the tightness measures shown in figure 3 is vacancies. This is the published JOLTS series from 2001:M1 to 2019:M12 and the Davis, Faberman, and Haltiwanger (2012a) backcast series for 1994:M1–2000:M12.[6] The denominators of the alternative generalized measures incorporate successively

6. We also have produced a labor market tightness measure using the Barnichon (2010) series for the pre-JOLTS period. We find results that are quite similar whether we use the Davis, Faberman, and Haltiwanger (2012a) or the Barnichon (2010) series.

Figure 3. Standard versus Generalized Measures of Labor Market Tightness (Constant Relative Search/Intensities, Recruiting Intensity = 1)



Sources: Authors' calculations using CPS, Job Openings and Labor Turnover Survey (JOLTS) for 2000:M12–2019:M12 and backcast vacancies from Davis, Faberman, and Haltiwanger (2012a) for 1994:M1–2000:M11 (DFH (2012a) vacancies hereafter).

Notes: All measures normalized to 1 in 2006. Std = V/U; Gen, U only = V/ES (unemployed only); Gen, U + Want = V/ES (unemployed plus want a job); Gen, U + OLF = V/ES (excludes employed); Gen, All = V/ES (all twenty- two groups).

more encompassing pools of effective searchers, in each case with the subgroups we have defined weighted in accord with their relative search intensities. The first generalized labor market tightness measure shown in figure 3 considers only the unemployed as effective searchers, while allowing for changes in unemployment composition. The second incorporates people who are out of the labor force but say they want a job, and the third adds the remainder of those out of the labor force. The final, fully generalized tightness measure also treats the employed as effective searchers. Once again, for ease of comparison, all of the measures in figure 3 have been normalized to equal 1 on average in 2006.

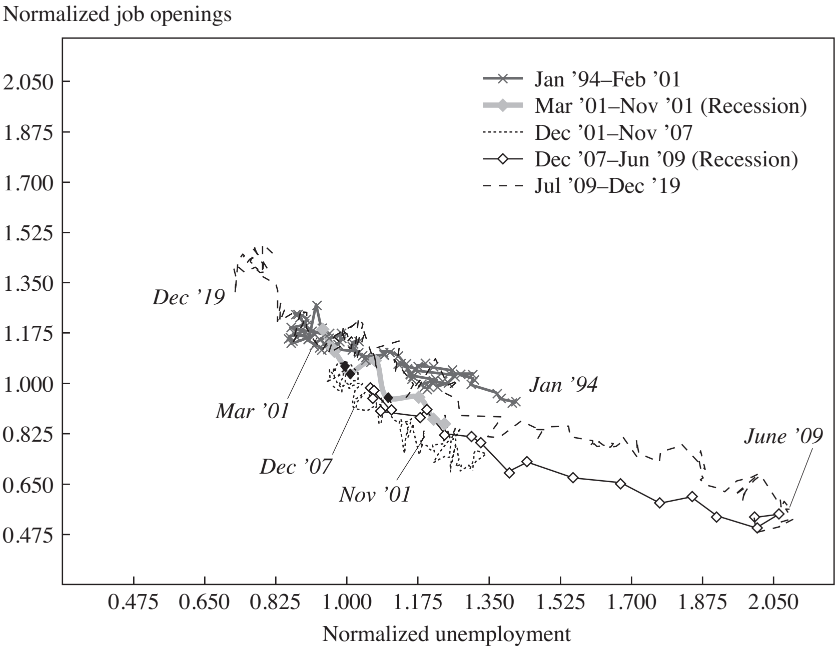
The more inclusive generalized tightness measures displayed in figure 3 are markedly less cyclical than the standard tightness measure—in particular, they fell much less steeply during the Great Recession and subsequently have risen less. The December 2019 values of the generalized measures incorporating only unemployed searchers or adding just the want a job group are not very different from the value of the standard measure, but the other measures are substantially lower, implying that the labor market was not as tight at that point as implied by the standard measure. To put this into context, the December 2019 value of the standard labor market tightness measure is about 42 percent higher than in March 2001. In contrast, the generalized measure using effective searchers is only about 17 percent higher. In short, our generalized measure using effective searchers in place of unemployment suggests a significantly different evolution of labor market tightness than the standard measure.

For comparison, online appendix figure A.4 plots labor market tightness measures constructed using the U6 index and the Richmond Fed index along with the standard and our generalized measure. The value of the tightness measure based on the U6 index is about 27 percent higher in December 2019 than it was in March 2001, and the value of the tightness measure based on the Richmond Fed index is about 18 percent higher.

IV. Beveridge Curve

A closely related but distinct way to look at the properties of the effective versus standard measures of searchers and job openings is through the lens of the Beveridge curve. Figure 4 displays the standard Beveridge curve using monthly data on vacancies and unemployment from 1994:M1 to 2019:M12. For this purpose, we use normalized unemployment and vacancy series defined relative to their 2006 average values. Plotting these series against one another makes clear their inverse relationship. In addition to the familiar downward sloping relationship between vacancies and unemployment, the figure also shows the substantial outward shift in that relationship during the long, slow recovery from the Great Recession. To illustrate, consider the period two to four years after the trough of the Great Recession (from June 2011 through May 2013), as compared to the comparable period following the trough of the 2001 recession (from November 2003 through October 2005). Job openings are only slightly lower over this portion of the recovery from the Great Recession than during the corresponding period following the 2001 recession (by about 8 percent) but unemployment is much higher (by about 48 percent). The marked increase in unemployment compared to that associated in the past with a similar level of vacancies led many to speculate that, following the Great Recession, there had been a decline in matching efficiency in the labor market.

Figure 4. Standard Beveridge Curve

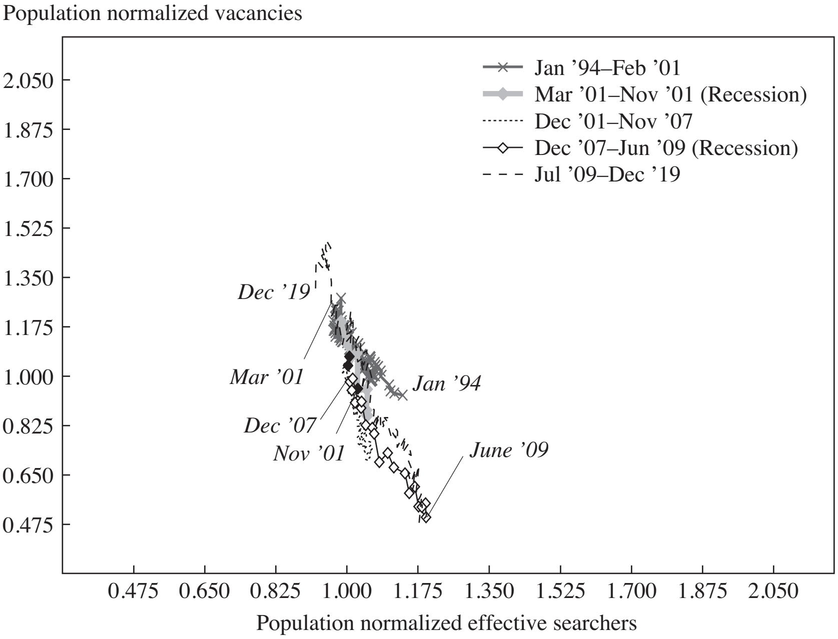


Source: Authors' calculations using job openings (vacancies) from JOLTS for 2000:M12–2019:M12 and DFH (2012a) vacancies for 1994:M1–2000:M11; unemployed from CPS.

Note: Both series as rates relative to population age 16+ normalized to 1 in 2006.

Figure 5 depicts the generalized Beveridge curve using effective searchers (using our baseline constant relative search intensities measure) with series normalized so the values plotted are all relative to their 2006 averages. The generalized Beveridge curve shown in figure 5 is much steeper than the standard version shown in figure 4. As noted previously when discussing figures 1 and 2, the proportional variation in effective searchers over time is much smaller than the proportional variation in unemployment. This translates into a normalized Beveridge curve that spans a much shorter distance along the horizontal axis than does the standard Beveridge curve and also is much more stable than the standard curve during the period following the Great Recession. Consider again the period two to four years after the trough of the Great Recession as compared to the period following the 2001 recession. As before, vacancies are slightly lower from two to four years after the Great Recession (by about

Figure 5. Beveridge Curve using Vacancies and Effective Searchers (Constant Relative Search Intensities, Recruiting Intensity = 1)



Source: Authors' calculations using job openings (vacancies) from JOLTS for 2000:M12–2019:M12 and DFH (2012a) vacancies for 1994:M1–2000:M11; effective searchers from CPS.

Notes: Both series as rates relative to population age 16+ normalized to 1 in 2006. Effective searchers as described in text.

8 percent), but whereas unemployment was *much* higher over the same period (about 48 percent), effective searchers were only slightly higher (about 10 percent).

One way to summarize the general shape of the Beveridge curve using different effective searcher measures is to fit descriptive regressions of vacancies on the various measures. Online appendix table B.3 shows that, consistent with figures 4 and 5, the standard Beveridge curve has a slope that is well below 1.0 in absolute value, whereas the Beveridge curve based on our fully generalized effective searcher measure is much steeper. The table also shows how moving in steps from the standard measure to our fully generalized measure leads to a Beveridge curve that is increasingly steep, reflecting the progressively lower volatility of the more encompassing measures. The slope of the Beveridge curve using the U6 measure is similar to that for the standard unemployment measure; the slope using the Richmond Fed index lies between that for the standard and the fully generalized measures.

V. Implications of Redefining Effective Searchers for the Matching Function

We have argued that the measure of labor market tightness using our generalized effective searcher series should be preferred conceptually to the standard measure, but we would like to have evidence that it actually does a better job of explaining the temporal variation in job-filling and job-finding rates. To evaluate the performance of the alternative measures in the matching function, we return to equation (1), the standard specification, and equation (7), the generalized specification, and ask how well each performs in tracking actual job-filling and job-finding rates. The targets we seek to match are the job-filling rate based on JOLTS data and the job-finding rate among the unemployed based on CPS gross flows data. We begin with our baseline model that defines $\rho_t^{V_i} = 1$ and $\rho_t^{S_i} = \gamma_i$. The next section of the paper will consider time-varying recruiting and relative search intensities.

V.A. Actual versus Model-Based Patterns in the Job-Filling Rate

Both the standard and the generalized matching function have implications for the evolution of the job-filling rate (H/V) as illustrated in equations (4) and (10). Because the left-hand sides of equations (4) and (10) are the same and are based on readily available data, we can compare the residual variation in the calibrated job-filling rates obtained using the standard and the generalized matching functions. For the present case, the specification of equation (10) is given by:

$$\text{(10')} \qquad \frac{H_t}{V_t} = \mu_t \left(\frac{\sum_i \gamma_i S_{it}}{V_t} \right)^{\alpha}.$$

In addition to our generalized measure with twenty-two different groups of effective searchers, we also consider versions of the generalized labor market tightness measure based on the U6 index and the Richmond Fed index.

In addition to the vacancy and unemployment (effective searcher) measures appearing in equations (4) and (10′), the calibrated job-filling rate associated with each of these tightness measures also depends on the elasticity of the matching function. To give each set of matching function arguments the best possible chance to fit the data well, we have estimated a separate α for each using a simplified version of the method proposed by Hall and Schulhofer-Wohl (2018). This method relates the job-finding rates for specific groups to vacancy duration from the JOLTS. As described more fully in online appendix B, this yields an estimate of the elasticity η of the job-finding rate with respect to vacancy duration for each measure of effective searchers. Given this estimate, we compute $\alpha = 1/(1 + \eta)$. For the standard model, this gives us an estimate for η of 1.04 (standard error 0.05) implying a value for α of 0.49. For the model with our effective searcher specification, the estimate is $\eta = 0.75$ (0.04) implying $\alpha = 0.57$. For the U6 measure, we obtain $\alpha = 0.48$ ($\eta = 1.1$ [0.05]) and for the Richmond Fed measure we obtain $\alpha = 0.60$ ($\eta = 0.67$ [0.05]). All of these estimates are reasonably similar and well within the middle of the range of estimates in the matching function literature (Petrongolo and Pissarides 2001). Our main results are broadly unchanged if we apply a common value of the matching function elasticity within the range of the separate estimates to calibrate the job-filling rates using the different tightness measures.[7]

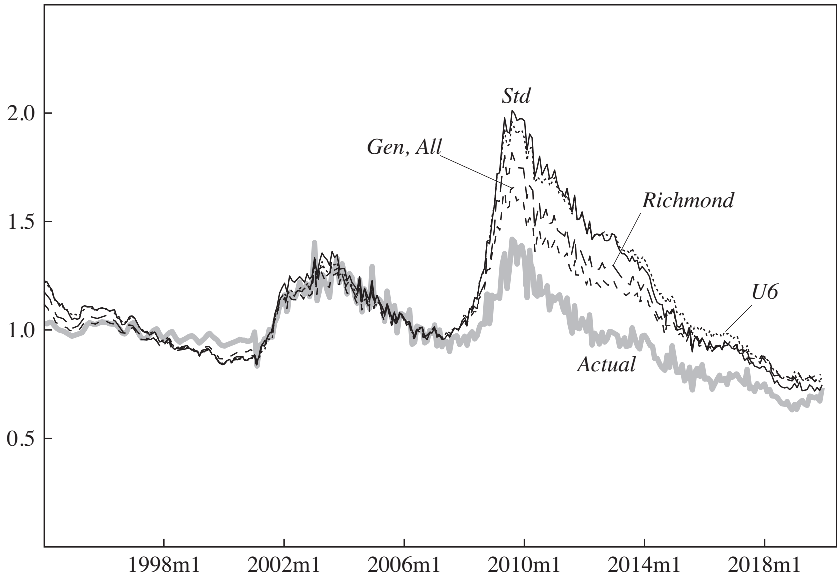
Figure 6 presents the actual and calibrated job-filling rates from equations (4) and (10′) using the standard, generalized, U6, and Richmond Fed measures. Once again, all series have been normalized to average 1 in 2006.[8] The calibrated job-filling rate based on equation (10′) and our generalized tightness measure tracks the actual job-filling rate much more closely than the calibrated rate based on equation (4) and the standard tightness measure. The U6 measure performs no better than the standard measure. The Richmond Fed measure performs substantially better than the standard measure, but not as well as our generalized measure.

To quantify the improvements in performance, panel A of table 3 reports the root-mean-square error (RMSE) of the calibrated job-filling rates as compared to the actual rates based on the different tightness measures. The generalized measure produces an RMSE of 0.13, a little more than half

7. In their evaluations of job-filling and job-finding rates using alternative measures of searchers and vacancies, Davis (2011), Davis, Faberman, and Haltiwanger (2013), and Mongey and Violante (2020) use $\alpha = 0.5$.

8. The model-specific normalization for the calibrated job-filling rates implies that we are permitting the mean matching efficiency for each model to be different. More specifically, it is permitted to vary in such a way that the calibrated job-filling rate in all models is equal to the actual job-filling rate on average in 2006.

Figure 6. Actual versus Calibrated Job-Filling Rates with Standard and Effective Searchers (Constant Relative Search Intensities, Recruiting Intensity = 1)



Source: Authors' calculations using CPS, JOLTS, DFH (2012a) vacancies, and Richmond Fed NEI.
Notes: All rates normalized to 1 in 2006. Actual = job filling rate (H/V) from JOLTS; Std = calibrated using V/U; Gen, All = calibrated using V/ES (all twenty-two groups); U6 = calibrated using V/U6 index; Richmond = calibrated using V/Richmond Fed NEI.

as large as the RMSE of 0.25 produced using the standard measure. In other words, using the generalized measure with effective searchers in the calibration substantially reduces the unexplained residual variation in the job-filling rate.[9]

For comparison purposes, we also show the RMSEs in the calibrated job-filling rates for the U6 and Richmond Fed indexes. The RMSE for the U6 index is identical to that for the standard measure; the RMSE for the Richmond Fed index is intermediate between those for the standard and the generalized model. Table 3 also reports summary statistics for the alternative calibrated job-filling rates compared to the actual. The calibrated

9. Note that the calibration exercise we carry out is quite different in nature from regressing the actual job-filling rate on the calibrated values. See online appendix B for further discussion.

Table 3. Relative Performance for Job-Filling and Job-Finding Rates using Standard versus Effective Searchers: Constant Relative Search Intensities, Recruiting Intensity = 1

	Std dev	*Corr w/ actual*	*RMSE*	*Ratio RMSE to standard*
A. Job-filling rate				
Actual	0.17	1.00	NA	NA
Standard	0.29	0.75	0.25	1.00
General, all	0.19	0.81	0.13	0.54
U6	0.27	0.70	0.25	1.00
Richmond Fed NEI	0.23	0.78	0.18	0.71
B. Job-filling rate using $\alpha = 0.57$ *for all measures*				
Standard	0.35	0.74	0.31	1.00
Gen, U only	0.31	0.79	0.25	0.81
Gen, U + want	0.28	0.78	0.23	0.75
Gen, U + OLF	0.21	0.78	0.16	0.53
Gen, all	0.19	0.81	0.13	0.43
C. Job-finding rate for the unemployed				
Actual	0.16	1.00	NA	NA
Standard ($\alpha = 0.49$)	0.20	0.31	0.26	1.00
Standard ($\alpha = 0.57$)	0.14	0.32	0.22	0.85
General, all	0.15	0.62	0.17	0.67

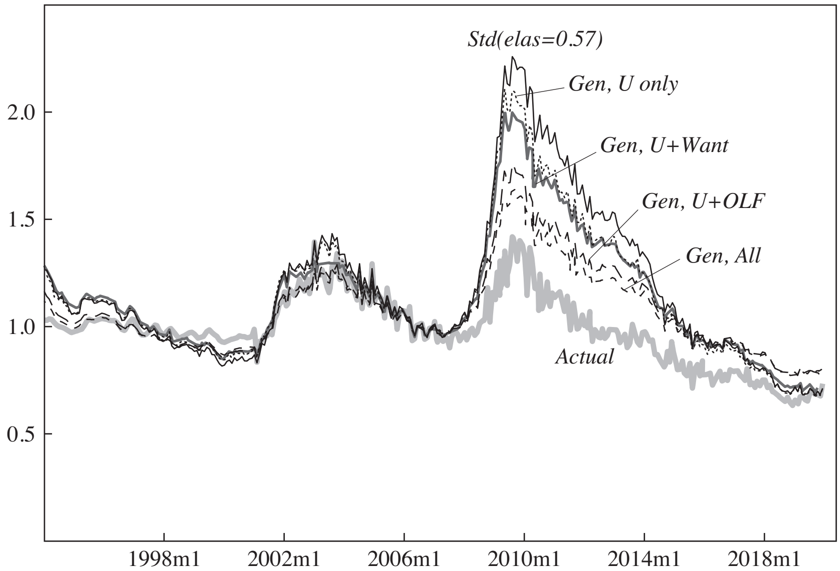
Source: Authors' calculations using CPS; JOLTS; Davis, Faberman, and Haltiwanger (2012a) vacancies; Richmond Fed NEI; and Bureau of Labor Statistics Gross Flows.

Notes: Panel A shows statistics for calibrated job-filling rates using matching function elasticities specific to each. Panel B shows statistics using same matching function elasticity for all measures. Panel C shows statistics for calibrated job-finding rate for the unemployed.

job-filling rate using effective searchers has a higher correlation and a standard deviation closer to the actual job-filling rate than the alternative calibrated series.

To help with understanding the factors underlying the improvement in performance of the generalized versus the standard tightness measure in tracking the job-filling rate, figure 7 presents calibrated rates based on a series of measures that incorporate the differences between the two in stages. For these figures and the associated analysis, we use the same matching function elasticity of $\alpha = 0.57$ for all of the counterfactual measures and for the standard measure. Panel B of table 3 reports the RMSEs in the calibrated job-filling rate using each of the different versions of the generalized tightness measure shown in figure 7. On its own, allowing for heterogeneity among the unemployed reduces the RMSE of the residual unexplained variation by about 20 percent. We gain an additional 5 percent by including in the pool of effective searchers those out of the labor force who want a job, an additional 20 percent by including the remaining people who are out of the labor force, and an additional 10 percent by including

Figure 7. Actual versus Calibrated Job-Filling Rates with Standard and Effective Searchers, Using $\alpha = 0.57$ for All Measures (Constant Relative Search Intensities, Recruiting Intensity = 1)



Source: Authors' calculations using CPS, JOLTS, and DFH (2012a) vacancies.
Notes: All rates normalized to 1 in 2006. Actual = job filling rate (H/V) from JOLTS; Std = calibrated using V/U; Gen, U only = calibrated using V/ES (unemployed only); Gen, U + Want = calibrated using V/ES (unemployed plus want a job); Gen, U + OLF = calibrated using V/ES (excludes employed); Gen, All = calibrated V/ES (all twenty-two groups).

the employed. Appropriate caution should be used in interpreting these figures, as the numbers we have reported are not an exact decomposition, but they do provide useful guidance with respect to which features of our generalized measure account for its better performance.

We also have explored a number of sensitivity checks that we summarize briefly here; details are shown in online appendix A. First, as shown in online appendix figure A.5, we replicate the analysis of job-filling rates using simple averages of the direct (raw) relative job-finding rates from table 1 rather than the relative job-finding rates that abstract from demographics. We also replicate the findings with a generalized measure that breaks the unemployed into just two groups, the short-term and the long-term unemployed, rather than thirteen more disaggregated groups (online appendix figure A.6) and with another version that disaggregates the population into only five groups (short-term unemployed, long-term unemployed, out of the labor force and want a job, out of the labor force and do not want a job, and employed; online appendix figure A.7). In both cases, the results are broadly similar to those we have just reported. The unexplained residual variation in the job-filling rate is the same using the raw job-finding rates and slightly higher when either limiting the disaggregation of the unemployed to just two groups or using the five-group disaggregation described above (RMSE = 0.14 for both as compared to RMSE = 0.13 for the fully disaggregated specification).

In addition, we have replicated the job-filling rate analysis using the Barnichon (2010) vacancy estimates based on help-wanted advertising for the 1994:M1–2000:M12 period in place of the series based on the methodology described by Davis, Faberman, and Haltiwanger (2012a). Again, the results are broadly similar (see online appendix figure A.8), though over the 1994:M1–2000:M12 period for which we must use projected vacancies, the generalized measure using Davis, Faberman, and Haltiwanger's (2012a) methodology performs better in tracking the job-filling rate than the Barnichon (2010) series. Over that period, the RMSE using the Davis, Faberman, and Haltiwanger series is 0.04 while the RMSE using the Barnichon series is 0.09.

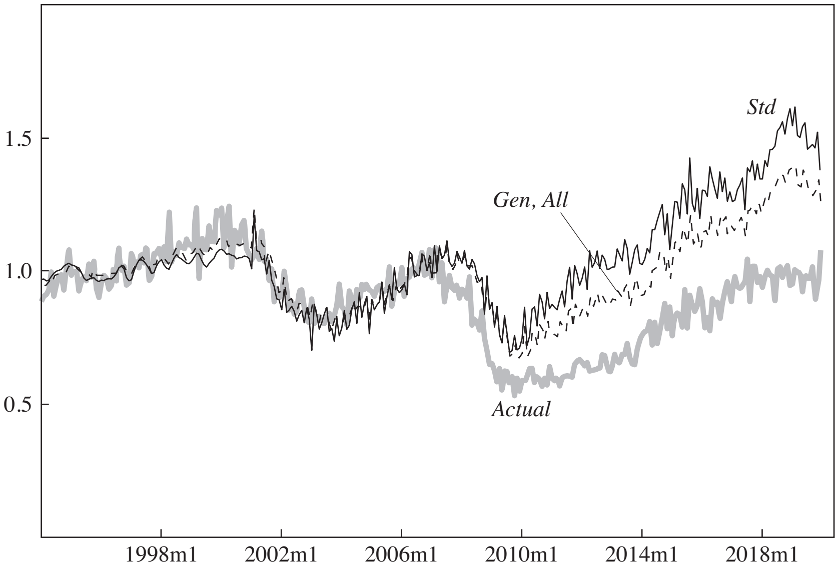
V.B. Actual versus Model-Based Patterns in the Job-Finding Rate

We now turn to investigating the performance of the generalized versus standard matching function for tracking job-finding rates. As noted in section II, overall hires per unemployed person is not a meaningful outcome measure for the generalized model with effective searchers. Instead, we use equations (14) and (15) to calibrate the job-finding rate of the unemployed. In the case we are currently considering, the specification of equation (14) is given by:

$$(14') \qquad \frac{H_{ut}}{U_t} = \frac{H_{ut}}{S_{ut}} = \mu_t^{1+\eta} (V_t/H_t)^{\eta} \left(\sum_{i \in u} \gamma_i S_{it}/S_{ut}\right).$$

Like the standard model of equation (15), the right-hand side of equation (14′) includes vacancy duration—vacancies relative to hires—but with the difference that this is now hires from all sources, not just hires from unemployment. This reflects the fact that, in the generalized model, the job-finding rate per effective searcher is assumed to be the same across all effective searcher groups. The generalized model also has an extra term that reflects the ratio of effective searchers among the unemployed to the number unemployed, with a higher ratio implying a larger number of hires per unemployed person.

Figure 8. Actual versus Calibrated Job-Finding Rates for the Unemployed with Standard and Effective Searchers (Constant Relative Search Intensities, Recruiting Intensity = 1)



Sources: Authors' calculations using CPS, JOLTS, DFH (2012a) vacancies, and BLS Gross Flows.
Notes: All rates normalized to 1 in 2006. Actual = job-finding rate for unemployed from published BLS Gross Flows; Std = calibrated using V/U with $\alpha = 0.49$ ($\eta = 1.04$); Gen, All = calibrated using V/ES (all twenty-two groups).

Figure 8 shows the actual and calibrated job-finding rates for the unemployed for the generalized and standard matching functions based on using equations (14′) and (15), respectively. As with our analysis of job-filling rates, we normalize both the actual and the calibrated series to be equal to 1 in 2006.[10] The actual job-finding rate is highly procyclical and falls especially sharply in the Great Recession. Both the generalized and the standard matching function track the job-finding rate among the unemployed reasonably well during the period prior to the Great Recession, though the generalized matching function performs somewhat better over that period. The generalized and the standard matching function track the sharp decline in the job-finding rate among the unemployed during the Great Recession about equally well, but the standard model implies a faster

10. As before, this implies that we are permitting the mean matching efficiency to differ across the standard and generalized models in such a way that both yield values in 2006 with mean equal to 1.

recovery whereas the calibrated generalized model lies closer to what actually happened.

Panel C of table 3 quantifies the improvement in performance in the calibrated job-finding rate for the unemployed from using the generalized rather than the standard matching function. The RMSE of the generalized calibration of the job-finding rate from unemployment is about a third lower than the RMSE for the standard calibration. As with the calibrated job-filling rate, the calibrated job-finding rate using effective searchers has a higher correlation and a standard deviation closer to the actual job-finding rate than the standard calibrated series.

Summary statistics reported in table 3 also shed light on the difference that the choice of α can make in the results obtained. Using $\alpha = 0.57$ (the elasticity for the generalized matching function) reduces the RMSE of the standard calibration by about 15 percent compared to its RMSE when using $\alpha = 0.49$. On the other hand, comparing the results in panels A and B of table 3, increasing α from 0.49 to 0.57 worsens the relative performance of the standard model for calibrating the job-filling rate. The implication of these conflicting effects is that the relatively poor performance of the standard model cannot be rescued with an alternative estimate of the matching function elasticity.

As already noted in comparing figure 1 with online appendix figures A.1, A.2, and A.3, the time series behavior of effective searchers is quite similar whether we use raw or demographically adjusted base period job-finding rates to weight the different groups of searchers; two groups or thirteen groups among the unemployed; or five broader groups of searchers rather than twenty-two groups as in our baseline analysis. In online appendix figures A.9, A.10, and A.11, we show results for the job-finding rate from unemployment for all three of these variants. The RMSE for the measure based on raw base period job-finding rates is nearly identical to that for our baseline specification. Both the RMSE for the generalized model with only two unemployment groups and the RMSE when we use just five groups rather than twenty-two groups are about 70 percent of that for the standard model, as compared to 67 percent for the fully general model.

VI. Exploring Time Variation in Recruiting Intensity and Search Intensity

The generalized model we outlined in section II includes both cross-sectional and time series variation in both recruiting intensities and search intensities, but thus far we have considered only the cross-sectional variation.

Research on alternative approaches to measuring the variation in recruiting intensity and relative search intensities is still in its early stages. In this section, we present results from an exploratory analysis of the variation in these intensities over our 1994:M1–2019:M12 sample period. For recruiting intensity, we use the Davis, Faberman, and Haltiwanger (2013) methodology to construct an aggregate recruiting intensity index. Using cross-sectional micro data, Davis, Faberman, and Haltiwanger (2013) estimate the elasticity between the job-filling rate and the gross hiring rate at the establishment level as $\phi = 0.82$. Applying this micro-based elasticity, they proxy aggregate recruiting intensity as $\rho_t^v = \left(\frac{H_t}{E_t}\right)^{\phi}$. We apply that proxy here.[11] It also is possible that, in addition to depending on the gross hiring rate, recruiting intensity may change over time due to composition effects (Davis, Faberman, and Haltiwanger 2012b) or to the endogenous responses of firms to overall labor market conditions, independent of their gross hiring rate (Gavazza, Mongey, and Violante 2018; Mongey and Violante 2020). We leave these possibilities for future exploration.

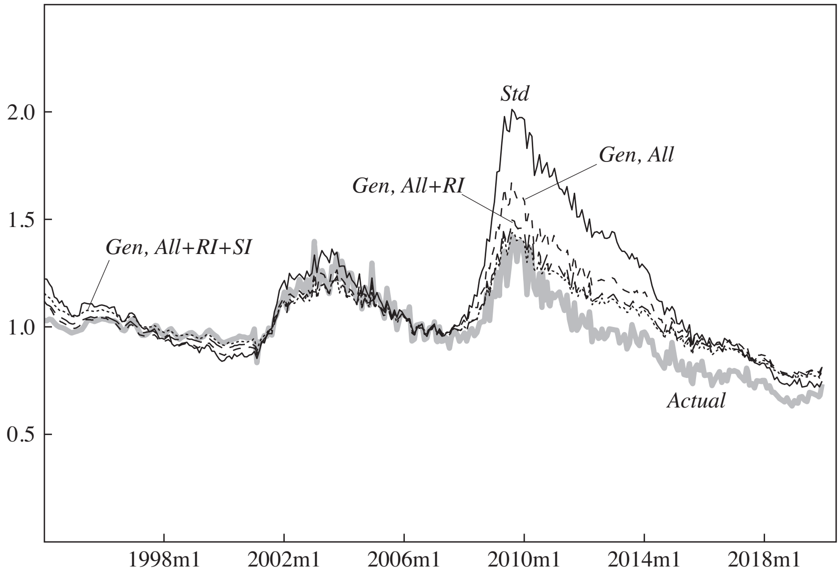
For time variation in relative job search intensities, we return to the specifications of Hall and Schulhofer-Wohl (2018). Their estimation allows not only for differences in average search intensities (what we have denoted as γ_i) but also for differences across job searcher types in the elasticity of job finding with respect to vacancy duration. We build on this insight to construct time-varying relative search intensities for our twenty-two groups, exploiting group-specific differences in the time variation in relative job-finding rates to infer what is happening to relative search intensities.[12] Details are reported in online appendix B. We do not use the common variation in job-finding rates in reweighting the different job searcher groups as we cannot distinguish changes in job-finding rates inherent in the matching function from those due to changes in search intensity.

Because the approach we have taken to constructing our time-varying recruiting intensity and relative job search intensity measures makes use of information related to the number of hires over time, some might be concerned that we have somehow built in the better performance of our model for explaining the job-filling and job-finding rates. Given the nature of our exercise, however, we do not believe this to be the case. We are

11. Online appendix figure A.12 shows how this recruiting intensity measure has moved over time.

12. See online appendix figure A.13 for the effective searcher measure using time-varying relative search intensities.

Figure 9. Actual versus Calibrated Job-Filling Rates with Standard and Generalized Tightness Measures including Time-Varying Recruiting (RI) and Search Intensity (SI)



Source: Authors' calculations using CPS, JOLTS, and DFH (2012a) vacancies.

Notes: All rates normalized to 1 in 2006. Actual = job filling rate (H/V) from JOLTS; Std = calibrated using V/U; Gen, All = calibrated using V/ES (all twenty-two groups) with constant relative search intensities and constant recruiting intensity; Gen, All + RI = calibrated using V/ES (all twenty-two groups) with constant relative search intensities and time-varying recruiting intensity; Gen, All + RI + SI = calibrated using V/ES (all twenty-two groups) with time-varying search intensities and recruiting intensity.

evaluating alternative functional forms for the matching function by comparing the calibrated job-filling and job-finding rates based on each to the actual data.[13] There is no inherent reason that incorporating time-varying recruiting and search intensity into these calibrated rates in fact will reduce the unexplained residual variation in the outcome of interest. Indeed, as will become clear, we obtain mixed results when we extend the general model in these directions.

Figure 9 depicts the actual and calibrated job-filling rates from the standard model, the general model with fixed relative job search intensities, the extension of that specification to include time-varying recruiting intensity, and finally a specification that includes both time-varying recruiting

13. Online appendix B includes the functional forms of job-filling and job-finding rates in this extended version of the general model.

Table 4. Relative Performance for Job-Filling and Job-Finding Rates using Standard versus General Model: Time-Varying Recruiting Intensity (RI) and Relative Job Search Intensity (SI)

	Std dev	*Corr w/ actual*	*RMSE*	*Ratio RMSE to standard*
A. Job-filling rate				
Actual	0.17	1.00	NA	NA
Standard	0.29	0.75	0.25	1.00
General, constant SI, RI = 1	0.19	0.81	0.13	0.54
General, time-varying RI	0.15	0.88	0.09	0.38
General, time-varying SI	0.17	0.84	0.12	0.49
General, time-varying RI and SI	0.14	0.91	0.09	0.34
B. Job-finding rate for the unemployed				
Actual	0.16	1.00	NA	NA
Standard	0.20	0.31	0.26	1.00
General, constant SI, RI = 1	0.15	0.62	0.17	0.67
General, time-varying RI	0.18	0.77	0.14	0.54
General, time-varying SI	0.29	0.68	0.22	0.85
General, time-varying RI and SI	0.31	0.75	0.22	0.83

Source: Authors' calculations using CPS; JOLTS; Davis, Faberman, and Haltiwanger (2012a) vacancies; and Bureau of Labor Statistics Gross Flows.

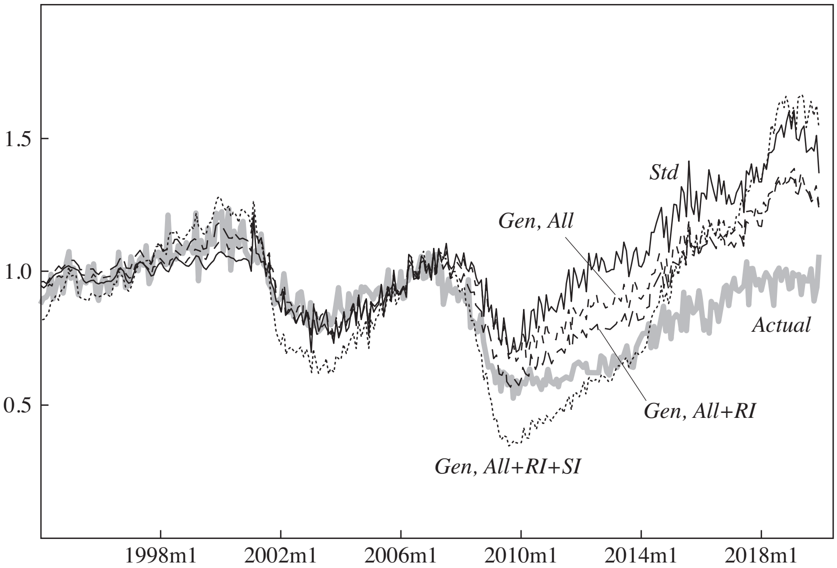
Notes: Panel A shows statistics for calibrated job-filling rates using matching function elasticities specific to each. Panel B shows statistics for calibrated job-finding rate for the unemployed.

intensity and time-varying relative search intensities. As can be seen in table 4, adding time-varying recruiting intensity helps to reduce the residual unexplained variation in the job-filling rate, lowering the RMSE in the generalized model's projected job-filling rate from 0.13 to 0.09. Allowing in addition for time-varying relative job search intensity has a much smaller effect that is not apparent when the numbers are rounded to two digits to the right of the decimal point, though the fourth column of the table shows that the RMSE declines relative to that in the standard model.

Figure 10 depicts the analogous exercise for the job-finding rate for the unemployed. Just adding time-varying recruiting intensity yields a modest improvement in the RMSE for the residual unexplained variation relative to the baseline case, lowering it from 0.17 to 0.14, but incorporating the time series variation in relative job search intensities leads to a worsening in the model's performance for tracking the job-finding rate among the unemployed, raising the RMSE to 0.22.

We regard the results shown in figures 9 and 10 as more suggestive than conclusive. The proxies we use for time variation in recruiting intensity and job search intensities are relatively crude and indirect. Developing better measures of these intensities is an important direction for future research.

Figure 10. Actual versus Calibrated Job-Finding Rates for the Unemployed with Standard and Generalized Tightness Measures including Time-Varying Recruiting (RI) and Search Intensity (SI)



Source: Authors' calculations using CPS, JOLTS, DFH (2012a) vacancies, and BLS Gross Flows.

Notes: All rates normalized to 1 in 2006. Actual = job-finding rate for unemployed from published BLS Gross Flows; Std = calibrated using V/U; Gen, All = calibrated using V/ES (all twenty-two groups) with constant relative search intensities and constant recruiting intensity; Gen, All + RI = calibrated using V/ES (all twenty-two groups) with constant relative search intensities and time-varying recruiting intensity; Gen, All + RI + SI = calibrated using V/ES (all twenty-two groups) with time-varying search intensities and recruiting intensity.

VII. Conclusions and Next Steps

The generalized measure of labor market tightness we have constructed based on the ratio of vacancies to effective searchers suggests that the US labor market was considerably less tight at the end of 2019 than implied by the standard ratio of vacancies to unemployment. The differing behavior of the two measures reflects the fact that the standard tightness measure does not account for important variation in search behavior on the part of workers. Job searchers include not only the unemployed but also those who are out of the labor force and the employed. In downturns, a more general index of effective searchers rises proportionally less than unemployment. The fact that the unemployed are only about 30 percent of all effective searchers contributes to this result, as any percentage increase in unemployment has a proportionally smaller effect on the overall number of effective searchers. Another contributing factor is that, during a protracted contraction such as the Great Recession, the distribution of unemployment shifts toward the long-term unemployed, meaning that effective searchers rise less than the simple unemployment count.

The central question motivating our analysis is whether substituting a generalized measure of effective searchers for the standard unemployment measure reduces the need to appeal to fluctuations in matching efficiency to explain what is happening to employment flows. We observe that the Beveridge curve constructed using effective searchers is much more stable than the standard Beveridge curve. Further, the matching function for hires with our generalized measure of labor market tightness as its argument outperforms the matching function based on the ratio of vacancies to unemployment. Specifically, the calibrated job-filling rate (hires per vacancy) using the generalized measure tracks the actual job-filling rate much more closely than the job-filling rate calibrated using the standard measure of labor market tightness. The calibrated job-finding rate among the unemployed (hires from unemployment per unemployed person) based on the generalized measure also comes closer to tracking the actual series than the calibrated rate based on the standard measure. These findings imply that our approach reduces the unexplained residual variation required in the matching function to be consistent with the real-world data.

Our baseline effective searcher measure is constructed using data for twenty-two separate population groups. We also have examined a number of alternatives that are less inclusive or based on more aggregated population groups. Taken together, the results make clear that the key to better matching the actual time series behavior of the job-filling and job-finding rates is having a broad-based measure of effective searchers that also distinguishes among core groups. Measures that include the employed and, especially, those out of the labor force do better than those limited to the unemployed, but the very detailed disaggregation of the broader groups we have adopted does little better than the simpler five-category breakout we examined as a sensitivity check. Given the large differences in base-period job-finding rates across the detailed categories combined in the five-category version of the generalized measure, it might seem surprising that the further breakouts we apply do not add more to the performance of the job-filling and job-finding models. Over our period, the changes in composition occurring within the five more aggregated groups are not large enough to make much difference, but this might not always be the case. As there is little cost to using more disaggregated groups to construct the generalized tightness measure, it seems preferable to us to do so.

Our baseline measure of labor market tightness undoubtedly could be improved and built upon. In a suggestive analysis, we find that incorporating proxies for time variation in relative job search intensities and also taking into account variation in recruiting intensity across employers further reduces the unexplained variation in job-filling rates. We find mixed evidence on using these proxies for explaining the fluctuations in job-finding rates among the unemployed. There clearly is more to be done to develop time series measures of search and recruiting intensity.

One topic that we have deliberately avoided but that is of critically important interest is whether and how the generalized labor market tightness approach could improve our understanding of wage and price pressures. It would be interesting to explore the estimation of Phillips curve–type relationships using a generalized measure of labor market tightness rather than the unemployment rate gap as the central explanatory variable. Even if it is true that labor market tightness is a better predictor of wage and price changes than the unemployment rate, however, there are other sources of instability in the Phillips curve relationship that seem likely to pose problems for the estimation of such relationships. Still, given that estimating and interpreting Phillips curves is an active area of research and highly relevant for policymakers, there would be value in exploring the role of generalized labor market tightness measures in this context.

ACKNOWLEDGMENTS We are grateful to Daniel Aaronson, Steven Davis, Jason Faberman, Robert Hall, Chris Nekarda, Ayşegül Şahin, Sam Schulhofer-Wohl, Jim Stock, Robert Valletta, and Justin Wolfers for helpful conversations and comments on an earlier draft; to colleagues at the University of Maryland for comments at an early stage of this research that helped to sharpen our thinking about these issues; and to participants in both the Federal Reserve System's June 2019 Conference on Monetary Policy Strategy, Tools and Communication Practices and the March 2020 Brookings Panel on Economic Activity meeting for their comments.

References

Aaronson, Daniel, Luojia Hu, Arian Seifoddini, and Daniel G. Sullivan. 2015. "Changing Labor Force Composition and the Natural Rate of Unemployment," Chicago Fed Letter 338.

Abraham, Katharine G. 2015. "Is Skill Mismatch Impeding U.S. Economic Recovery?" *ILR Review* 68, no. 2: 291–313.

Abraham, Katharine G., John C. Haltiwanger, Kristin Sandusky, and James R. Spletzer. 2019. "The Consequences of Long-Term Unemployment: Evidence from Linked Survey and Administrative Data." *ILR Review* 72, no. 2: 266–99.

Ahn, Hie Joo, and Ling Shao. 2017. "Precautionary On-the-Job Search over the Business Cycle." Finance and Economics Discussion Series 2017–025. Washington: Board of Governors of the Federal Reserve System.

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

Barnichon, Regis, and Geert Mesters. 2018. "On the Demographic Adjustment of Unemployment." *Review of Economics and Statistics* 100, no. 2: 219–31.

Black, Matthew. 1980. "Pecuniary Implications of On-the-Job Search and Quit Activity." *Review of Economics and Statistics* 62, no. 2: 222–29.

Blanchard, Olivier, and Peter A. Diamond. 1989. "The Beveridge Curve." *Brookings Papers on Economic Activity*, no. 1, 1–76.

Blanchard, Olivier, and Peter A. Diamond. 1992. "The Flow Approach to Labor Markets." *American Economic Review* 82, no. 2: 354–59.

Blau, David M., and Philip K. Robins. 1990. "Job Search Outcomes for the Employed and Unemployed." *Journal of Political Economy* 98, no. 3: 637–55.

Broersma, Lourens, and Jan C. van Ours. 1999. "Job Searchers, Job Matching and the Elasticity of Matching." *Labour Economics* 6:77–93.

Casselman, Ben. 2018. "As Labor Pool Shrinks, Prison Time Is Less of a Hurdle." *New York Times*, January 13.

Crump, Richard K., Stefano Eusepi, Marc Gionnoni, and Ayşegül Şahin. 2019. "A Unified Approach to Measuring U*." *Brookings Papers on Economic Activity*, Spring, 143–214.

Davis, Steven J. 2011. Comment on "Job Search, Emotional Well-Being, and Job Finding in a Period of Mass Unemployment: Evidence from High-Frequency Longitudinal Data," by Alan Krueger and Andreas Mueller. *Brookings Papers on Economic Activity*, Spring, 58–70.

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger. 2012a. "Labor Market Flows in the Cross Section and Over Time." *Journal of Monetary Economics* 59, no. 1: 1–18.

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger. 2012b. "Recruiting Intensity during and after the Great Recession." *American Economic Review* 102, no. 3: 584–88.

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger. 2013. "The Establishment-Level Behavior of Vacancies and Hiring," *Quarterly Journal of Economics*, 128, no. 2: 581–622.

DeLoach, Stephen B., and Mark Kurt. 2013. "Discouraging Workers: Estimating the Impacts of Macroeconomic Shocks on the Search Intensity of the Unemployed." *Journal of Labor Research* 34:433–54.

Diamond, Peter A. 1982. "Wage Determination and Efficiency in Search Equilibrium." *Review of Economic Studies* 49, no. 2: 217–27.

Elsby, Michael, Ryan Michaels, and David Ratner. 2015. "The Beveridge Curve: A Survey." *Journal of Economic Literature* 53, no. 3: 571–630.

Faberman, R. Jason, Andreas I. Mueller, Ayşegül Şahin, and Giorgio Topa. 2017. "Job Search Behavior among the Employed and the Non-Employed." Working Paper 23731. Cambridge, Mass.: National Bureau of Economic Research.

Faberman, R. Jason, Andreas I. Mueller, Ayşegül Şahin, and Giorgio Topa. 2020. "The Shadow Margins of Labor Market Slack." Working Paper 26852. Cambridge, Mass.: National Bureau of Economic Research.

Fujita, Shigeru, and Guiseppe Moscarini. 2017. "Recall and Unemployment." *American Economic Review* 107, no. 12: 3875–916.

Fujita, Shigeru, Guiseppe Moscarini, and Fabien Postel-Vinay. 2019. "Measuring Employer-to-Employer Reallocation." Working Paper.

Gavazza, Alessandro, Simon Mongey, and Giovanni L. Violante. 2018. "Aggregate Recruiting Intensity." *American Economic Review* 108, no. 8: 2088–127.

Gomme, Paul, and Damba Lkhagvasuren. 2015. "Worker Search Effort as an Amplification Mechanism." *Journal of Monetary Economics* 75:106–22.

Hall, Robert, and Sam Schulhofer-Wohl. 2018. "Measuring Job-Finding Rates and Matching Efficiency with Heterogeneous Job-Seekers." *American Economic Journal: Macroeconomics* 10, no. 1: 1–32.

Haltiwanger, John C., Henry R. Hyatt, Lisa B. Kahn, and Erica McEntarfer. 2018. "Cyclical Job Ladders by Firm Size and Firm Wage." *American Economic Journal: Macroeconomics* 10, no. 2: 52–85.

Haltiwanger, John C., Henry R. Hyatt, and Erica McEntarfer. 2018. "Who Moves Up the Job Ladder?" *Journal of Labor Economics* 36, no. S1: S301–36.

Hornstein, Andreas, and Marianna Kudlyak. 2016. "Estimating Matching Efficiency with Variable Search Effort." Working Paper 2016–24. San Francisco: Federal Reserve Bank of San Francisco.

Hornstein, Andreas, Marianna Kudlyak, and Fabian Lange. 2014. "Measuring Resource Utilization in the Labor Market." *Federal Reserve Bank of Richmond Economic Quarterly* 100, no. 1: 1–21.

Jones, Stephen R. G., and W. Craig Riddell. 1999. "The Measurement of Unemployment: An Empirical Approach." *Econometrica* 67, no. 1: 147–62.

Kaitz, Hyman. 1970. "Analyzing the Length of Spells of Unemployment." *Monthly Labor Review* 93, no. 11: 11–20.

Katz, Lawrence F. 1986. "Layoffs, Recalls and the Duration of Unemployment." Working Paper 1825. Cambridge, Mass.: National Bureau of Economic Research.

Katz, Lawrence F., and Bruce D. Meyer. 1990. "The Impact of the Potential Duration of Unemployment Benefits on the Duration of Unemployment." *Journal of Public Economics* 41, no. 1: 45–72.

Krueger, Alan B., Judd Cramer, and David Cho. 2014. "Are the Long-Term Unemployed on the Margins of the Labor Market?" *Brookings Papers on Economic Activity*, Spring, 229–80.

Krueger, Alan B., and Andreas Mueller. 2010. "Job Search and Unemployment Insurance: New Evidence from Time Use Data." *Journal of Public Economics* 94:298–307.

Krueger, Alan B., and Andreas Mueller. 2011. "Job Search, Emotional Well-Being, and Job Finding in a Period of Mass Unemployment: Evidence from High-Frequency Longitudinal Data." *Brookings Papers on Economic Activity,* Spring, 1–57.

Kudlyak, Marianna. 2017. "Measuring Labor Utilization: The Non-Employment Index." Federal Reserve Bank of San Francisco Economic Letter 2017–08, March 17.

Modestino, Alicia Sasser, Danial Shoag, and Joshua Balance. Forthcoming. "Upskilling: Do Employers Demand Greater Skill when Workers Are Plentiful?" *Review of Economics and Statistics*.

Mongey, Simon, and Gianlucia Violante. 2020. "Macro Recruiting Intensity from Micro Data." Working Paper 26231. Cambridge, Mass.: National Bureau of Economic Research.

Mortensen, Dale T., and Christopher A. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *Review of Economic Studies* 61, no. 3: 397–415.

Mukoyama, Toshihiko, Christina Patterson, and Ayşegül Şahin. 2018. "Job Search Behavior over the Business Cycle." *American Economic Journal: Macroeconomics* 10, no. 1: 190–215.

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

Petrongolo, Barbara, and Christopher A. Pissarides. 2001. "Looking into the Black Box: A Survey of the Matching Function." *Journal of Economic Literature* 39, no. 2: 390–431.

Pissarides, Christopher. 2000. *Equilibrium Unemployment Theory*, 2nd ed. Cambridge, Mass.: MIT Press.

Şahin, Ayşegül, Joseph Song, Giorgio Topa, and Giovanni L. Violante. 2014. "Mismatch Unemployment." *American Economic Review* 104, no. 11: 3529–64.

Sedlacek, Petr. 2016. "The Aggregate Matching Function and Job Search from Employment and Out of the Labor Force." *Review of Economic Dynamics* 21:16–28.

Shimer, Robert. 2001. “The Impact of Young Workers on the Aggregate Labor Market.” *Quarterly Journal of Economics* 116, no. 3: 969–1008.

Shimer, Robert. 2004. “Search Intensity.” Working paper. Chicago: University of Chicago.

Smialek, Jeanna. 2019. “The Economy Is Strong and Inflation Is Low. That’s What Worries the Fed.” *New York Times*, May 21.

Veracierto, Marcelo. 2011. “Worker Flows and Matching Efficiency.” *Economic Perspectives* 35, no. 4: 147–69.

Comments and Discussion

COMMENT BY

AYŞEGÜL ŞAHIN Abraham, Haltiwanger, and Rendell developed a generalized measure of labor market tightness which takes into account intensive and extensive margins of search activity on both demand and supply sides of the labor market. Their measure captures the hiring process in the US economy better than the standard measure of labor market tightness. Their success is probably not surprising given that they build on two well-documented facts about the hiring process in the United States. The first is that the majority of jobs are filled by workers who are not unemployed. The second is that the number of vacancies is an imperfect proxy for firms' total recruiting effort since firms vary their recruiting intensity over the business cycle.

Assessment of labor market conditions is a fundamental issue in macroeconomics and a key input to implementation of monetary and fiscal policy. The authors' work is a valuable addition to the wealth of labor market indicators developed in the last decade to better evaluate labor market developments in light of ongoing secular trends. With the unemployment rate jumping from 3.5 percent in February 2020 to 14.7 percent in April 2020, the labor market will be our main focus of attention for years to come. Understanding how workers search for employment opportunities, how firms fill their open positions, and how search activity responds to macroeconomic conditions will help us in characterizing the adjustment path of the US labor market to the COVID-19 shock. This comment reviews and interprets the authors' findings and suggests new directions of research.

FRAMEWORK Abraham, Haltiwanger, and Rendell use the matching function specification—a key building block of the Diamond-Mortensen-Pissarides search and matching framework—to characterize the behavior of hiring in the US economy. In its basic form, the matching function takes the Cobb-Douglas form and specifies hires, h_t, as a function of two inputs: vacancies posted by firms looking to hire, v_t, and unemployed workers looking for jobs, u_t:

$$h_t = \Phi_t (u_t)^{\alpha} (v_t)^{1-\alpha}, \tag{1}$$

where $\alpha \in (0,1)$ is the unemployment share and Φ_t is the aggregate matching efficiency parameter. The job-finding and the job-filling rates can be written as a function of labor market tightness ($\theta_t = v_t/u_t$) as $\Phi_t\ \theta_t^{1-\alpha}$ and $\Phi_t (1/\theta_t)^{\alpha}$, respectively. This specification implies that θ_t is the only determinant of the hiring process if Φ_t does not vary over time. However, as discussed by the authors, this specification ignores the fact that most hires originate from employment or nonparticipation. Moreover, it does not take into account the intensive margins of firm and worker search effort. A generalized tightness measure, $\tilde{\theta}_t$, which incorporates these factors can be defined as

$$\tilde{\theta}_t = \frac{\sum_j \rho_t^{V_j} V_{jt}}{\sum_i \rho_t^{S_i} S_{it}}, \tag{2}$$

where $\rho_t^{V_j}$ is the recruiting intensity of vacancy group in firm type j and $\rho_t^{S_i}$ is the search intensity of searcher group i. The interpretation of $\tilde{\theta}_t$ is straightforward: it is the ratio of the recruiting-intensity-weighted sum of vacancies to the search-effort-weighted sum of searchers in the economy. While searchers include all individuals regardless of their labor force status, not all groups contribute equally to the total search effort in the economy. The contribution of each group depends on their relative search intensity, $\rho_t^{S_i}$.

Abraham, Haltiwanger, and Rendell do not explore the variation in the composition of firms by age, size, or industry over the business cycle. Instead, they use the direct measure of aggregate recruiting intensity from Davis, Faberman, and Haltiwanger (2013). With the aggregate recruiting intensity ρ_t^V in hand, they compute the effective vacancies as $\rho_t^V V_t$. The implementation of the generalized measure of effective searchers is more involved since it requires using detailed micro data from the Current Population Survey (CPS). First, the authors identify twenty-two distinct labor force states following Hall and Schulhofer-Wohl (2018) and compute the job-finding rates of these groups using the matched CPS data. They assume that the relative job-finding rate of each group corresponds to its relative search intensity. This assumption allows them to compute an aggregate measure of effective searchers by weighting each group i by its relative job-finding rate. Among these twenty-two detailed labor market groups, thirteen groups are unemployed; seven groups are those who are out of the labor force, and two are the employed. Table 1 of the paper shows that these groups vary considerably in their job-finding rates with recently temporarily laid-off workers having the highest job-finding rate.

ARE TWENTY-TWO LABOR MARKET STATES NECESSARY? Abraham, Haltiwanger, and Rendell differentiate between various detailed labor market states and end up using twenty-two distinct groups in their implementation. While this level of disaggregation allows them to capture even small variations in the relative job-finding rates of searchers, it restricts their analysis to the post-1994 period since the questions that allow them to differentiate between these twenty-two states are not available in the CPS before 1994. I show that focusing only on a few broad labor market states is sufficient to capture almost all the variation in the measure of effective searchers over time. This simplification is useful for two reasons. First, it makes it possible to compute the generalized tightness measure without using the CPS micro data. Second, it allows us to compute the generalized tightness measure starting in 1948.

Starting with five broad labor force states, the measure of effective searchers can be simplified as:

$$S_t = \underbrace{\rho^{UST}U_t^{ST} + \rho^{ULT}U_t^{LT}}_{\text{Unemployed}} + \underbrace{\rho^{W}W_t + \rho^{N}N_t}_{\text{Nonparticipants}} + \underbrace{\rho^{E}E_t}_{\text{Employed}}.$$

I focus on unemployment, nonparticipation, and employment and further distinguish between short-term (U^{ST}) and long-term unemployed (U^{LT}) among the unemployed and between those who want a job (W) and those who do not (N) among nonparticipants. I approximate the relative job-finding rates of these groups using the raw job-finding rates from the authors' table 1 and choose the following weights:

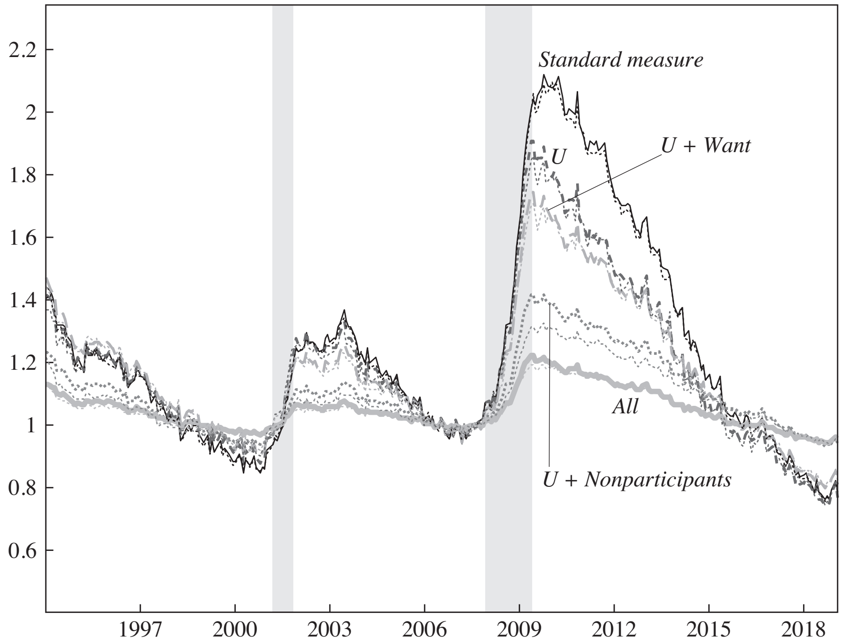
$$\rho^{UST} = 1, \rho^{ULT} = 0.48, \rho^{W} = 0.40, \rho^{N} = 0.09, \rho^{E} = 0.07.$$

I then compute the rate of effective searchers as

$$s_t = \frac{S_t}{P_t} = \frac{S_t}{U_t + N_t + E_t}$$

and normalize it to 1 in 2006. Figure 1 shows that the series computed using only five labor market states line up remarkably well with the authors'

Figure 1. Measures of Effective Searchers using Five Labor Force States



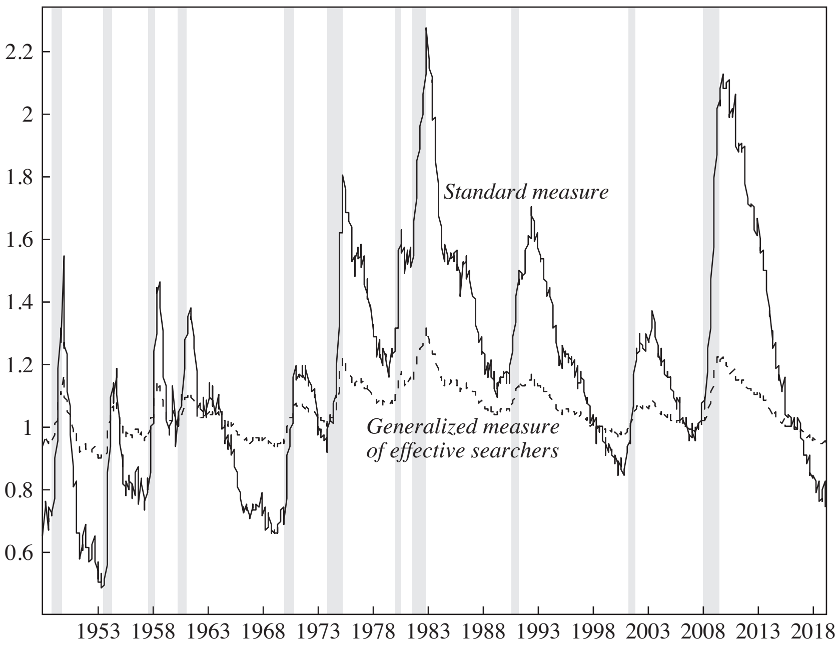
Source: Author's calculations based on the Current Population Survey (CPS).
Notes: AHR measures indicated by thin dotted lines. Standard measure: U_t/P_t; U: $(U_t^{ST} + 0.48U^{LT})/P_t$; U + Want: $(U_t^{ST} + 0.48U^{LT} + 0.40W_t)/P_t$; U + Nonparticipants: $(U_t^{ST} + 0.48U^{LT} + 0.40W_t + 0.09N_t)/P_t$; and all: $(U_t^{ST} + 0.48U^{LT} + 0.40W_t + 0.09N_t + 0.07E_t)/P_t$. Series normalized to 1 in 2006.

measures.[1] The intuition is clear: while job-finding rates vary substantially by detailed labor market state, most of these groups are too small to affect the aggregate measure. Therefore I conclude that focusing on five broad labor market states does not change the essence of the authors' findings but simplifies their analysis considerably by making it possible to compute the index without using the CPS micro data.

HISTORICAL PERSPECTIVE Given that the five-state simplification does remarkably well, I now turn to historical data and calculate the measure of effective searchers starting in 1948. This requires consolidating the two groups of nonparticipants to only one, but even this abstraction does not change the behavior of the generalized measure. Figure 2 plots the generalized measure of effective searchers along with the standard measure

1. While these weights are not exact, following my discussion at the panel, the authors repeated their analysis with only five groups and found very similar results to mine.

Figure 2. Generalized Measure of Effective Searchers and the Standard Measure in the 1948–2018 Period



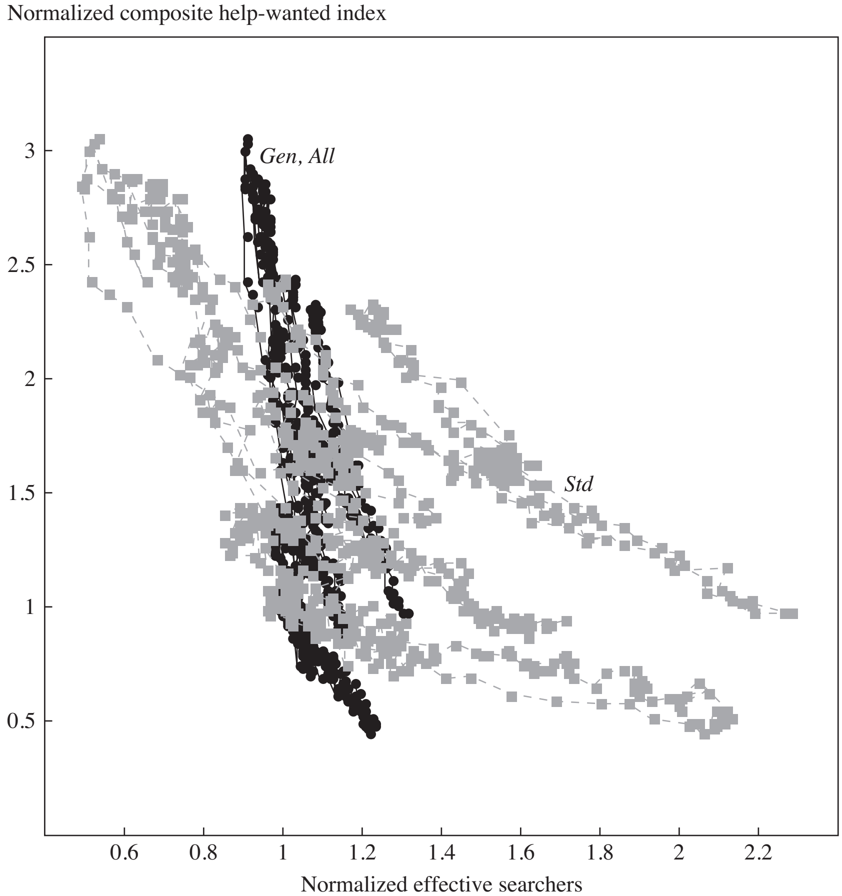
Source: Author's calculations based on the Current Population Survey (CPS).
Notes: AHR measures indicated by dotted lines. Standard measure: U_t/P_t; generalized measure of effective searchers: $(U_t^{ST} + 0.48U^{LT} + 0.09N_t + 0.07E_t)/P_t$.

starting in 1948 and shows that the deviation between the two measures is not specific only to the most recent expansion. The effective searchers measure exhibits more muted countercyclical behavior than the standard measure throughout the sample—a distinct feature of the generalized measure that I will discuss below in detail.

Interestingly, despite differences in the magnitude of their cyclicality, measures based on unemployment and effective searchers provide a very similar assessment of the historical business cycles. Both measures imply that the 1973–1975, 1980–1982, and 2007–2009 recessions were the deepest downturns and the second halves of the 1960s, 1990s, and 2010s were the tightest labor markets of the postwar period.

I also examine the historical Beveridge curve using the standard measure and the measure of effective searchers combined with Barnichon's (2010) composite help-wanted vacancy index in figure 3. The Beveridge

Figure 3. Historical Beveridge Curves using the Generalized Measure of Effective Searchers and the Standard Measure

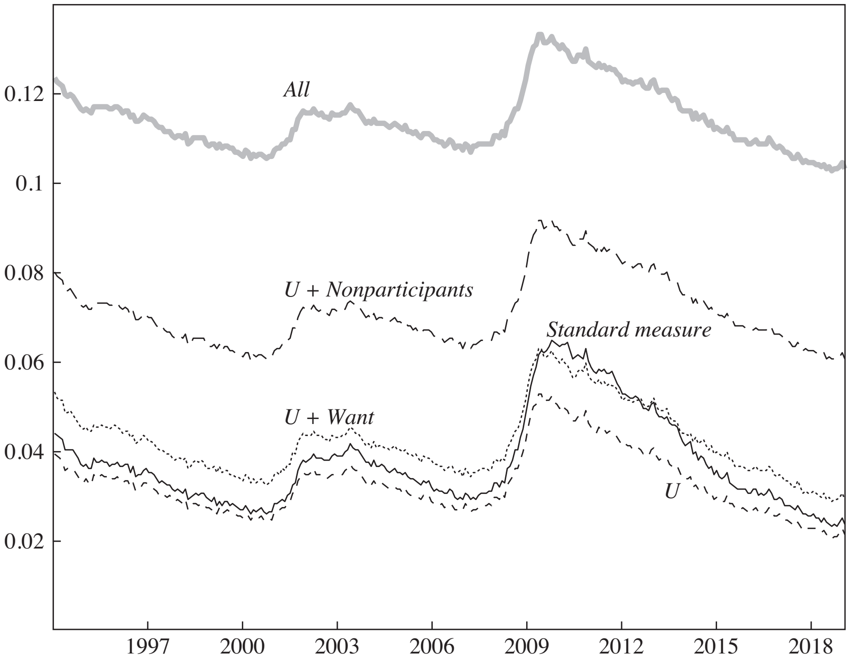


Source: Author's calculations based on the Current Population Survey (CPS) and composite help-wanted vacancy index developed by Barnichon (2010).

Note: Historical Beveridge curve using the unemployment-to-population ratio (gray dots) and the effective measure of searchers (black dots).

curve constructed using unemployment exhibits substantial horizontal shifts over time as is well documented in Diamond and Şahin (2015). The shifts are much less pronounced when the measure of effective searchers is used. However, the Beveridge curve also becomes much steeper when it is constructed with the new measure. The steepness of the Beveridge curve suggests that fluctuations in vacancies have very little effect on labor

Figure 4. Measures of Effective Searchers in Levels



Source: Author's calculations based on the Current Population Survey (CPS).

Notes: Alternative searcher measures in levels. Standard measure: U_t/P_t; U: $(U_t^{ST} + 0.48U^{LT})/P_t$; U + Want: $(U_t^{ST} + 0.48U^{LT} + 0.40W_t)/P_t$; U + Nonparticipants: $(U_t^{ST} + 0.48U^{LT} + 0.40W_t + 0.09N_t)/P_t$; and all: $(U_t^{ST} + 0.48U^{LT} + 0.40W_t + 0.09N_t + 0.07E_t)/P_t$.

market underutilization. This interesting observation deserves further discussion and analysis since it suggests that the link between labor demand and labor underutilization has been considerably weaker than is typically assumed in the postwar period.

WHY IS THE COUNTERCYCLICALITY DAMPENED? Abraham, Haltiwanger, and Rendell focus only on normalized measures of effective searchers and do not report or interpret the *levels* of their measures. While this type of normalization is useful in the context of the matching function framework, I argue that the difference in the levels of alternative measures of searchers is important for the interpretation of the authors' findings.

Figure 4 plots different measures of searchers in levels without normalizing them to 1 in 2006. Comparison of figures 1 and 4 shows that measures plotted in levels look very different than their normalized counterparts. Expanding the pool of searchers by adding employed workers and nonparticipants increases the level substantially despite the low relative job-finding rates of these groups, and the normalization reduces cyclicality of the measures substantially.

This comparison is very helpful in understanding the reasons for dampened countercyclicality of the effective searchers measure relative to unemployment. First, the level of the generalized measure is higher than the level of the standard measure due to the addition of the high number of employed workers and nonparticipants. Second, there are offsetting changes in the composition of searchers over the business cycle. During recessions, unemployment increases but due to the decline in employment, the number of employed searchers declines, dampening the rise in unemployment. Similarly, during expansions as the number of unemployed searchers declines, the number of employed searchers increases, moderating the decline in the number of unemployed searchers. As a result, the generalized measure of effective searchers fluctuates less than unemployment over the business cycle. This finding has an important implication about the hiring incentives of firms: firms do not find it that much easier to fill jobs during recessions despite high unemployment since the number of employed searchers declines. This is why the generalized tightness measure does better in capturing the behavior of hires in the matching function framework.

HOW SHOULD WE ASSESS LABOR MARKET CONDITIONS? The Great Recession and the subsequent period of sluggish recovery in the labor market triggered an important line of research, with numerous studies developing labor market indicators as alternatives to the unemployment rate. Most of these measures exploit one of the following two approaches. The first focuses on estimating a time-varying natural rate of unemployment and uses the unemployment gap—the difference between the actual and the natural rate of unemployment—as a measure of labor market tightness. While this line of research typically focuses on estimating the natural rate of unemployment using only simple aggregate data, more recent work unifies the macro approach with rich labor market data (Crump and others 2019). The main advantage of this approach is that it provides a unified framework that takes into account secular demographic trends, wage and price inflation, and inflation expectations. Therefore it directly connects to maximum employment and price stability objectives.

The second approach is to develop broader measures which take into account additional margins of labor market underutilization. A common practice is to weight different groups of workers depending on their demographic characteristics, wages, search activity, or job-finding rates, such as in Perry (1970), Hornstein, Kudlyak, and Lange (2014), and this paper. While implementing this approach, choice of weights is an important issue since the interpretation of the aggregate measure depends heavily on the weights used. For example, the authors choose to use relative job-finding rates and show that their measure relying on realized transitions is informative about the hiring process in the economy. However, this does not necessarily mean that their measure would be preferable to other measures for other purposes. For example, one fundamental issue in macroeconomics is to estimate the potential output of the economy. In that case, an alternative measure which weights workers by their average idle hours (desired hours minus actual hours worked), as in Faberman and others (2020), would be informative about potential output.

To conclude, the authors' generalized tightness measure is a valuable addition to our arsenal of labor market indicators. They convincingly demonstrated that the measure they develop captures the essence of the hiring process well. Future work calls for broadening the metric for success by considering the usefulness of the new metric in capturing wage growth and inflation. This extension would help to connect the new measure more directly to monetary policy implementation. With the unemployment rate at its highest level in the postwar period, we need as much information as possible about the labor market for real-time assessment of the state of the labor market and its medium-term evolution. The work of Abraham, Haltiwanger, and Rendell undoubtedly will contribute to our understanding of the labor market in the years to come.

REFERENCES FOR THE ŞAHIN COMMENT

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

Crump, Richard K., Stefano Eusepi, Marc Giannoni, and Ayşegül Şahin. 2019. "A Unified Approach to Measuring U*." *Brookings Papers on Economic Activity*, Spring, 143–214.

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger. 2013. "The Establishment-Level Behavior of Vacancies and Hiring." *Quarterly Journal of Economics* 128, no. 2: 581–622.

Diamond, Peter A., and Ayşegül Şahin. 2015. "Shifts in the Beveridge Curve." In *Research in Economics*, edited by Federico Etro. Amsterdam: North-Holland.

Faberman, R. Jason, Andreas I. Mueller, Ayşegül Şahin, and Giorgio Topa. 2020. "The Shadow Margins of Labor Market Slack." Working Paper 26852. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26852.

Hall, Robert, and Sam Schulhofer-Wohl. 2018. "Measuring Job-Finding Rates and Matching Efficiency with Heterogeneous Job-Seekers." *American Economic Journal: Macroeconomics* 10, no. 1: 1–32.

Hornstein, Andreas, Marianna Kudlyak, and Fabian Lange. 2014. "Measuring Resource Utilization in the Labor Market." *Federal Reserve Bank of Richmond Economic Quarterly* 100, no. 1: 1–21.

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

COMMENT BY

JUSTIN WOLFERS This paper by Abraham, Haltiwanger, and Rendell follows a long tradition of Brookings papers in proposing a new measure of labor market tightness. The new measure aims to capture effective job search effort, and its key advantages are that it recognizes that the unemployed aren't the only group who search and it accounts for the reality that different groups of workers search with different degrees of success. To be precise, the proposed new measure—which the authors call the "generalized measure of effective searchers"—is constructed as a weighted average of the share of the population in twenty-two different labor market states, with the weights reflecting each group's baseline level of search effectiveness (measured by each group's historical job-finding rate). As a weighted average rather than simple average, it generalizes more standard metrics like unemployment (which effectively puts a weight of one on the unemployed and zero on all others).

This paper was written during what now seems a distant time in early 2020 when the official unemployment rate was lower than it had been in about half a century and one of the central macro policy questions of the day was whether unemployment could go any lower. The final draft was submitted a few weeks later, when the coronavirus pandemic had pushed the unemployment rate to levels higher than at any time since the Great Depression. A superficial assessment might argue that a paper titled "How Tight Is the US Labor Market?" has been overtaken by events which have rendered it less relevant. But in reality, this research question may actually be more relevant than ever before, as the recession caused by the coronavirus—which includes elements of demand shocks and supply shocks—has led to arguably greater uncertainty about the extent of slack in the labor market than the United States has experienced at any point throughout the postwar period. The present-day relevance simply requires a reframing from this being a paper assessing how best to measure labor market tightness to one assessing how best to measure labor market looseness.

In proposing a new measure of labor market conditions, this paper enters an already quite crowded marketplace. Between the various measures of labor market slack published by the Bureau of Labor Statistics (BLS), alternatives tracked by the Federal Reserve banks, the preferred metrics of an array of private-sector forecasters, and new measures proposed in back issues of *Brookings Papers on Economic Activity*, it's no exaggeration to say that dozens of measures have been proposed and are closely tracked. Whether it's worth tracking one more measure depends on whether this new measure yields additional useful information.

The main test that the paper offers is to ask whether this new measure does a better job in explaining time series variation in the job-filling and job-finding rates than standard metrics like the official BLS measure of unemployment. The authors argue that their measure does add useful additional information. The analysis that I present below disagrees. To preview, I find little evidence that their proposed new measure outperforms standard measures. While the authors claim the generalized measure yields a better fit, the tests reported reflect arbitrary auxiliary assumptions that tilt the playing field against the standard measure of unemployment. When these assumptions are relaxed, the proposed new measures no longer appear to be more predictive. Indeed, the proposed new measure appears to largely track the official measure of unemployment. By this telling, there is no problem with the newly proposed measure, but there is also little to recommend it, as it adds little information beyond that in the standard measure.

EXPLAINING TIME SERIES VARIATION IN THE JOB-FILLING RATE The main test that the paper implements is to ask which measure of labor market slack—the proposed generalized measure of effective searchers or the widely used official measure of unemployment—does a better job explaining time series variation in the job-filling rate, which is the ratio of hiring to vacancies. To motivate their analysis, the authors start with a Cobb-Douglas matching function of the form:

$$H_t = \mu U_t^{\alpha} V_t^{1-\alpha}.$$

This can be understood as a production function, where the output is the number of people hired, H_t, and the production of these matches is a function of the aggregate effort workers put into searching for jobs (U_t) and the aggregate effort that firms put into searching for workers (V_t). In addition, μ is an index of the efficiency of this production function, and the empirical analysis in this paper holds this parameter constant, which is why it lacks a time subscript.

It is common for empirical analyses of the matching function to use the level of unemployment (which I'll denote U_t^s) as a proxy for total search effort by workers (so $U_t = U_t^s$) and the level of vacancies as a proxy for total search effort by firms (V_t). The authors propose that their generalized measure of effective searchers (which I'll denote U_t^g) may be a better proxy for the level of total search effort by workers (so $U_t = U_t^g$, instead). In a later extension, they also argue that accounting for recruiting intensity may yield a better proxy of the search effort by firms.

To test this, the paper asks which competing proxy measure of U_t—the standard measure, which is the number of unemployed people U_t^s, or their generalized measure of effective searchers U_t^g—better explains the observed time series movements in the job-filling rate. The specific empirical exercise it implements comes from a simple rearrangement of the matching function so that the dependent variable is now the job-filling rate:

$$\frac{H_t}{V_t} = \mu\left[\frac{V_t}{U_t}\right]^{-\alpha}.$$

The variable to be forecast, H_t/V_t, is readily measured as monthly hires divided by monthly vacancies, both from the Job Openings and Labor Turnover Survey (JOLTS) data. Likewise, the two candidate inputs into this forecasting equation—U_t/V_t, measured as either U_t^S/V_t, the ratio of the unemployed (from the Current Population Survey [CPS]) to the number of vacancies, or as U_g^g/V_t, the ratio of the generalized measure of effective searchers (constructed by the authors from CPS data) to the number of vacancies—are easily measured.

So far, this sounds like a standard forecast evaluation exercise. But there's a twist. No statistical agency publishes estimates of either μ (the efficiency of the matching function) or α (the Cobb-Douglas coefficient in the matching function). Think of the parameter μ as determining the position of the Beveridge curve, while the parameter α determines its slope. The values of these parameters will shape how well any measure of slack will fit the data. The values that are imposed reflect an auxiliary set of assumptions, and as I will show below, these assumptions largely drive the empirical findings that follow.

Unfortunately, the paper does not contain much discussion of where these parameters come from nor much of a defense of the specific parameter values that are imposed. In order to appropriately assess the quality of the evidence presented in the paper, it's worth being more transparent about these assumptions. First, μ reflects an arbitrary normalization, in which the time series of forecasts of the job-filling rate (and the inputs into that forecast) are indexed to be equal to 1, on average, during 2006. This effectively imposes

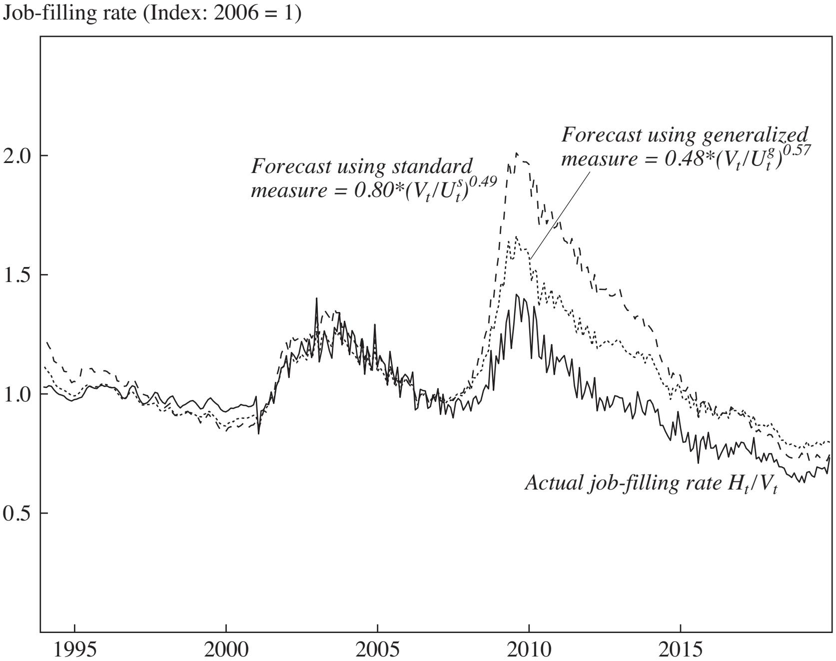
$$\mu = \frac{\overline{H_t/V_t}^{2006}}{\overline{(V_t/U_t)^{-\alpha}}^{2006}},$$

which would make sense if μ were constant and the matching function exactly fit the data (with an error term equal to zero) on average through 2006. When U_t is proxied by the level of unemployment, this yields $\widehat{\mu^s} = 0.80$, but when it is proxied by the generalized measure of effective searchers, it yields $\widehat{\mu^g} = 0.48$. Next, α comes from a structural estimation procedure described only in online appendix B of the paper, in which the evolution over time of the job-finding rates for each of the twenty-two groups are estimated as a function of a group fixed effect, a common time trend, and the aggregate vacancy-to-hires ratio. In this framework, α is recovered as a nonlinear function of the coefficient on the vacancy-to-hires ratio, and the various α's for each of the twenty-two groups are averaged to get an aggregate $\widehat{\alpha^g}$. (I append the superscript *g* to denote that this is the α estimated for the *g*eneralized measure). However, this equation is not estimated directly but rather using instrumental variables, where employment from the payroll survey is an instrument for the vacancy-to-hires ratio. For the generalized measure of effective searchers, this yields $\widehat{\alpha^g} = 0.57$. For the standard measure, this procedure collapses so that there is only one group of searchers (the unemployed), and so the dependent variable is the ratio of unemployment-to-employment flows to unemployment, and the dependent variables are a constant, a time trend, and the aggregate vacancy-to-hires ratio (which again is instrumented using payroll employment). This yields $\widehat{\alpha^s} = 0.49$ (where the superscript *s* denotes that this is the α that applies to the *s*tandard measure).

Imposing these assumptions yields the forecasts of the job-filling rate shown in figure 1 (which replicates exactly the corresponding lines from figure 5 in the paper). The top (dashed) line is the forecast generated using the standard measure of U_t, the next (dotted) line is the forecast generated using the generalized measure, while the actual job-filling rate, which is shown as a solid line, lies below both of these forecasts for most of the sample. The authors emphasize that the forecast generated using the standard

Figure 1. Predicting the Job-Filling Rate

Using **imposed** matching function coefficients



Source: Author's calculations, from data in the paper.

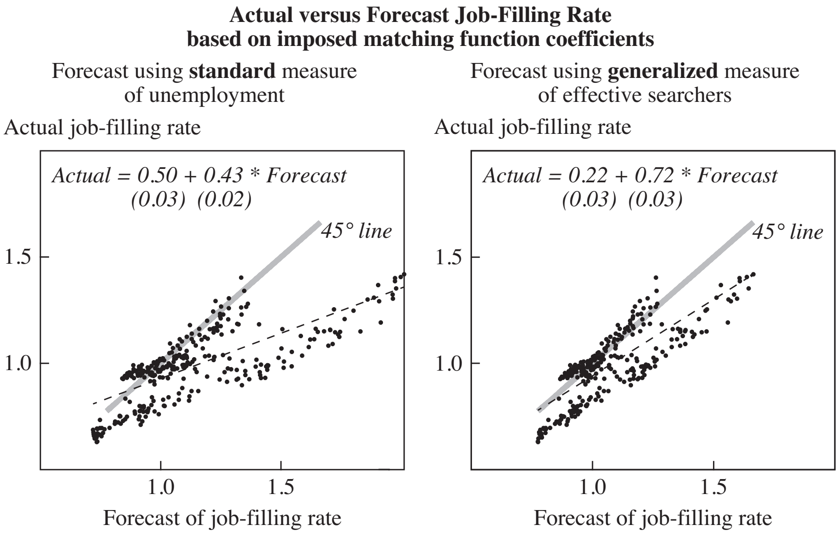
measure is further from the actual outcomes than the forecast generated using the generalized measure. This difference is the basis of the claim that the standard measure is outperformed by the authors' preferred generalized measure.

But eyeballing this figure reveals something more troubling: both lines look like problematic forecasts. Both are (substantially) higher on average than the actual job-filling rate. And both forecasts rise and fall substantially more over the business cycle than the actual job-filling rate. Neither forecast appears to fit the data well.

This is easy to confirm using standard forecast evaluation methods. In particular, figure 2 shows scatter plots of the actual job-filling rate against each of these forecasts. Neither fits the data well. Figure 2 also reports the results of regressions of the following form:

$$Actual_t = a + b \times Forecast_t + \epsilon_t.$$

Figure 2. Forecast Evaluation



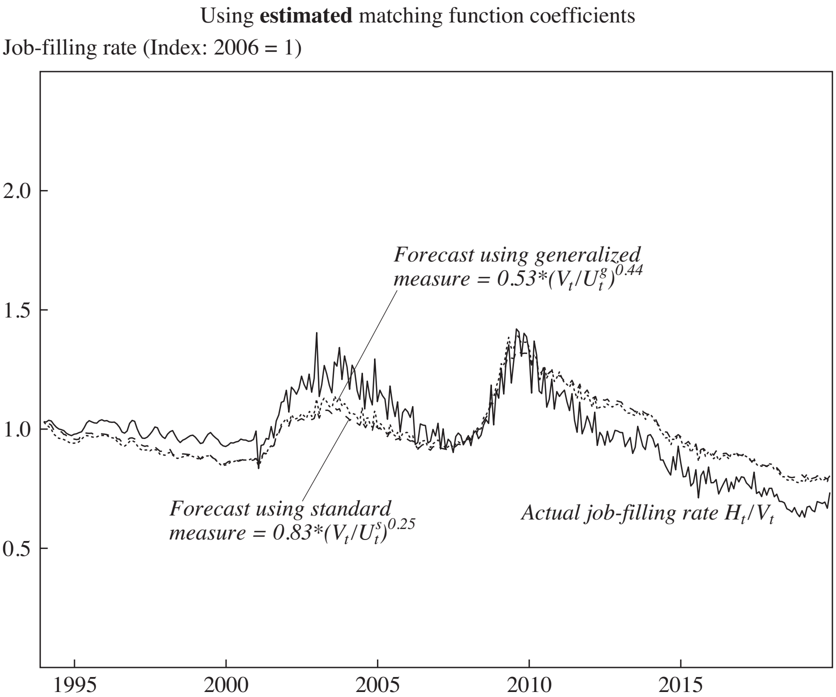
Source: Author's calculations.

An unbiased and efficient forecast would yield a constant term of zero and a coefficient on the forecast of one. (Failing this test implies that forecast errors are forecastable, and hence statistically inefficient.) The left panel shows that the forecast generated using the standard measure of unemployment fails both of these tests, and the right panel shows that the forecast generated using the generalized measure of effective searchers also fails the same test. (F-tests of the joint null that $a = 0$ and $b = 1$ are overwhelmingly rejected in both cases.)

These artifacts reflect the fact that the μ and α coefficients that were imposed appear to be inappropriate if one is interested in forecasting the job-filling rate. In addition, as we shall see in a moment, the differences between the two forecasts are largely attributable to these imposed coefficients. Correcting for these artifacts reveals that both the standard and the generalized measure yield very similar forecasts.

A more transparent (and arguably less arbitrary) approach to figuring out what values of μ and α to apply for this forecast evaluation exercise would be to simply estimate each of these parameters directly and then use those estimates to assess the accuracy of the ensuing forecasts. This alternative would be more consistent with the authors' claim that their analysis aims to estimate α in a way that would "give each set of matching function

Figure 3. Predicting the Job-Filling Rate Based on Estimated Matching Function Coefficients



Source: Author's calculations.

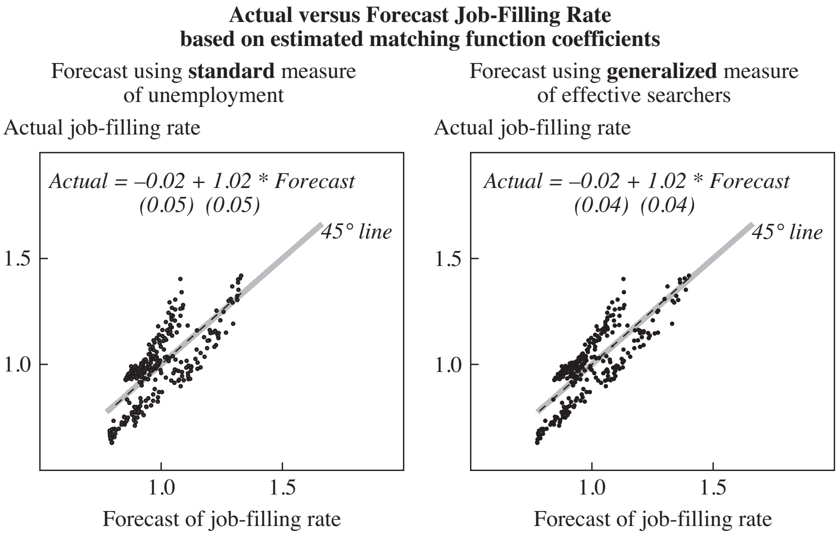
arguments the best possible chance to fit the data well." As such, I ran the following simple nonlinear least squares regression:

$$\frac{H_t}{V_t} = \mu\left[\frac{V_t}{U_t}\right]^{-\alpha}.$$

Initially, I ran this regression using the standard measure of U_t^s (the number of unemployed people), which yielded estimates of $\widehat{\mu^s} = 0.83$ (se = 0.01) and $\widehat{\alpha^s} = 0.25$ (se = 0.01). Running this regression, but using the generalized measure of effective searchers U_t^s as the measure of worker search instead, yielded estimates of $\widehat{\mu^g} = 0.53$ (se = 0.014) and $\widehat{\alpha^g} = 0.44$ (se = 0.02). (Running the regression in logs: $\log(H_t) = \log(\mu) + \alpha\log(U_t) + (1-\alpha)\log(V_t)$ yielded relatively similar estimates of α in each case.)

Figure 3, which is drawn on the same scale as figure 1, plots each of the resulting forecasts of the job-filling rate, along with the actual job-filling

Figure 4. Forecast Evaluation Based on New (Estimated) Forecasts



Source: Author's calculations.

rate. This yields a dramatically different picture than figure 1 (or figure 5 in the paper under discussion). Both sets of forecasts of the job-filling rate are now much more accurate.

Most importantly, the dashed and the dotted lines are now almost identical. Once one abandons the arbitrarily imposed matching function coefficients, the forecast for the job-filling rate generated using the standard measure of unemployment is almost identical to that forecast using the generalized measure of effective searchers. Indeed, the correlation between these two series is 0.991.

This analysis yields an interpretation of the paper's findings that is largely at odds with the interpretation offered by the authors. It suggests that any differences in the performance of these measures is not due to richer information embedded in the generalized measure of effective searchers but rather is due to the specific coefficients imposed on the matching function in generating forecasts. As such, the authors' claim that "the relatively poor performance of the standard model cannot be rescued with an alternative estimate of the matching function elasticity" appears to be wrong.

Not surprisingly, estimating the coefficients that we use to generate forecasts of the job-filling rate yields better behaved forecasts. Figure 4 plots each of our new forecasts of the job-filling rate against the actual

Table 1. Formal Forecast Evaluation

	Standard deviation	*Correlation with actual*	*RMSE*	*Ratio of RMSE to standard*
Panel A: Generating forecasts using imposed coefficients (from table 3 of the paper)				
Actual job-filling rate	0.17	1.00	—	—
Forecast using standard measure of unemployment	0.29	0.75	0.25	1.00
Forecast using generalized measure of search activity	0.19	0.81	0.13	0.54
Panel B: Generating forecasts using estimated coefficients (preferred alternative)				
Actual job-filling rate	0.17	1.00		
Forecast using standard measure of unemployment	0.12	0.76	0.11	1.00
Forecast using generalized measure of search activity	0.13	0.81	0.10	0.90

Source: Author's calculations.

job-filling rate based on the coefficients estimated above. Forecasts generated using either the standard or the generalized measure of labor market slack are unbiased, and they both rise and fall in proportion with the business cycle. Of course, this occurs largely because these forecasts are generated using coefficients that were estimated with an eye to closely matching the actual outcomes. (In the regression of actual outcomes on forecasts, while the constant is close to zero, it is not precisely zero—and the slope is close to one, but not precisely one—because these forecasts were generated from an equation that is not linear in the parameters.)

At this point, it is worth revisiting the authors' evaluation of the relative performance of each measure. Table 1 presents a formal forecast evaluation. Panel A replicates the findings the authors present in their table 3, showing that using the imposed matching function coefficients leads to the conclusion that the root-mean-square error (RMSE) of forecasts generated using the standard measure is much higher (that is, worse) than that of the forecasts generated using the generalized measure of search activity. Indeed, the authors' preferred measure yields forecast errors that tend to be nearly half that when using the standard measure. A Diebold-Mariano test reveals that this difference in RMSE is (highly!) statistically significant.

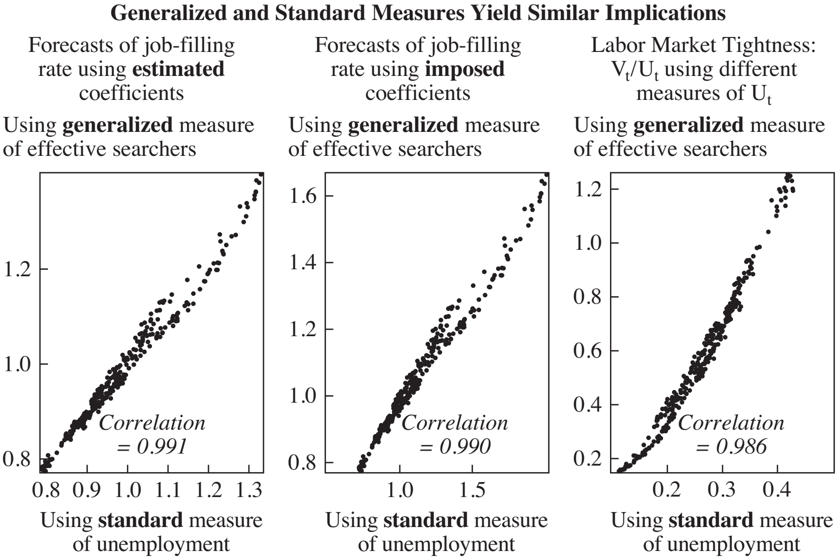
But this finding no longer survives when one is no longer willing to impose the specific values that the authors impose on the matching function coefficients. To see this, panel B has a parallel structure to panel A, but it reports on the accuracy of the forecasts generated using each measure of labor market slack, but using the matching function coefficients that I estimated above. The fit of both sets of forecasts—generated using either measure of labor market slack—is now much better. Moreover, the performance of the forecasts generated using the competing measures of slack—whether measured as the correlation between the forecast and the actual outcome or as the RMSE—are now remarkably similar. The RMSE of the two alternative forecasts evaluated in panel B are no longer economically meaningful, and a Diebold-Mariano test fails to reject the null that they're equal. All told, the evidence in panel B is that there's little to recommend one measure over the other. In turn, this reflects the reality that the alternative forecasts are remarkably similar.

THE REMARKABLE SIMILARITY OF ALTERNATIVE MEASURES OF SLACK The most striking finding so far is that using the estimated matching function coefficients yields forecasts of the job-filling rate that are almost identical whether using the standard measure of unemployment or the generalized measure of effective searchers as an input. It's worth pausing a bit to see where this comes from.

The left panel of figure 5 plots the forecast of the job-filling rate generated using the generalized measure of effective searchers versus the forecast generated using the standard measure of unemployment. This panel uses my estimated matching function coefficients, and as reported above, it reveals a correlation between these measures of 0.991. The middle panel turns to a similar plot, but this time the forecasts are based on the matching function coefficients imposed by the authors. The correlation remains very high. This suggests that the high correlation of the forecasts generated by the two competing measures of labor market slack is not driven by my preference for estimating the matching function coefficients rather than imposing them, as this high correlation between these two sets of forecasts is a feature even in the authors' preferred measures. The third panel shows the source of this similarity, plotting the basic measure of labor market tightness, V_t/U_t, when estimated using the two alternative measures of U_t. The measure of slack based on the generalized measure of effective searchers, V_t/U_t^s, is remarkably similar to that generated using the standard measure of unemployment, V_t/U_t^s.

This, in turn, implies that the differences in the forecasts of these two measures aren't driven by one measure being more informative than the other. Both rise and fall almost in tandem. Rather, they are driven by difference in the scale and location of the two measures of labor market slack. Indeed, this is the key to reconciling the evidence in figure 1 which shows that the authors' forecasts of the job-filling rate based on their generalized

Figure 5. Comparing Generalized and Standard Measures



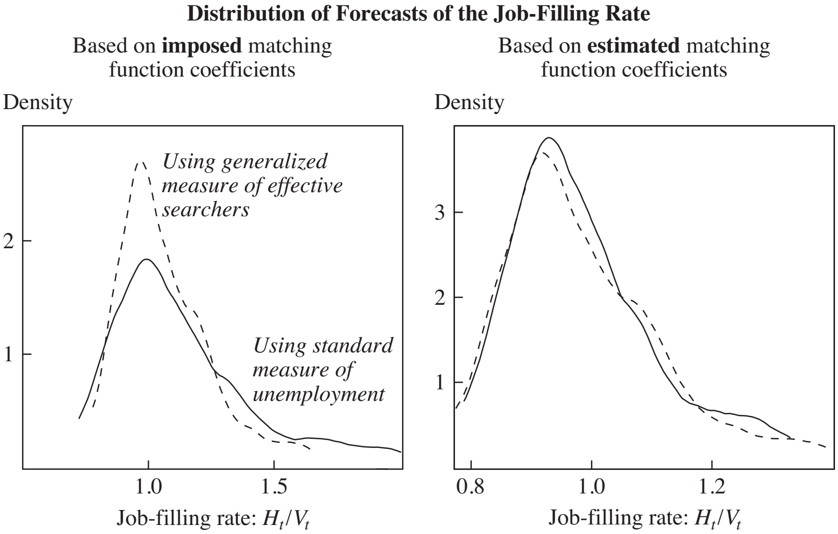
Source: Author's calculations.

measure of effective searchers is quite different from their forecasts based on the standard measure of unemployment, with the evidence in figure 5 that these two sets of forecasts are still very highly correlated. As figure 1 shows, both measures rise and fall in lockstep, but using the coefficients imposed by the authors yields a series of forecasts of the job-filling rate based on the standard measure that are typically located higher than those based on the generalized measure, and they are scaled so that they are more variable.

The location and scale of each series is determined by the coefficients α and μ. Thus, nearly *all* of the difference in the forecasts generated by these series reflects the choices of these auxiliary parameters. Remarkably little of the difference is due to the generalized measure bringing extra information about the state of the labor market.

Figure 6 illustrates the role that the matching function coefficients play in determining the scale and location of each series. It should be read in concert with the right panel of figure 5, which illustrates that the inputs into these forecasts are remarkably highly correlated. Yet as the left panel of figure 6 shows, the matching function coefficients imposed by the authors yields a series of forecasts based on the standard measure of unemployment

Figure 6. Scale and Location of Alternative Forecasts of the Job-Filling Rate



Source: Author's calculations.

that is both larger on average and much more variable than the forecasts based on the generalized measure of effective searchers. The right panel shows that these differences in scale and location are no longer meaningful when using estimated matching function coefficients.

DO THE IMPOSED COEFFICIENTS MAKE SENSE? At this point, it is worth assessing whether there is any reason to prefer the tests based on the imposed matching function coefficients presented in the paper versus the estimated coefficients emphasized here. On this score, there are two key points to make, one methodological and the other quantitative.

First, the methodological point yields a prima facie case that the matching function coefficients imposed in the paper are not appropriate for this sort of analysis. The exercise the paper performs is effectively a forecast evaluation exercise. But the source of the imposed α's comes from an instrumental variables regression, designed to recover the "structural" matching function coefficient. While structural coefficients are useful for constructing policy counterfactuals, in the presence of measurement error (and this paper is motivated by the idea that there is measurement error in our measures of labor market slack) a structural coefficient will not generally yield an efficient or unbiased forecast. (Recall, while ordinary least squares may yield biased coefficients, those coefficients are still the best linear unbiased estimates.) Indeed, figure 3 illustrates that the imposed coefficients yield forecasts of the job-filling rate that are both biased and inefficient. Even if policymakers were to rely on the generalized measure of effective searchers, it is hard to believe that they would rely on these imposed coefficients to evaluate the likely implications for the job-finding rate.

To further muddy the methodological waters, the μ coefficients that the authors impose are neither structural nor reduced form, but the result of an arbitrary normalization in which it is essentially assumed that the unemployment and vacancy data for 2006 determine the location of the Beveridge curve. As far as I can tell, the authors never defend this choice, but it turns out not to be an innocuous normalization, as it also helps determine how strongly the job-filling rate responds to changes in labor market tightness.

Second, there is a quantitative point to be made. The authors are correct to argue that they need to use different estimates of α and μ when generating forecasts based on the two different measures of slack. Different estimates are needed because the generalized measure of effective searchers is scaled in a way that does not permit direct comparison to the standard measure, which is the level of unemployment. (The generalized measure is the sum over twenty-two population groups of the product of their share in the population at that point in time and their relative job-finding rate in 2006. Its precise scale depends on the normalization of these job-finding rates, or γ_i, in the authors' terminology.) As such, it makes sense to use different estimates of α_s and α_g to generate forecasts based on either the standard or generalized measures of labor market tightness.

But while the precise numeric values of α_s and α_g used to generate each forecast may differ, for this to be an apples-to-apples comparison, they should have a similar *economic* interpretation. They should each predict a similar response of the job-filling rate (H_t/V_t) to business cycle fluctuations in labor market tightness ($\theta_t = V_t/U_t$). After all, only by ensuring that both comparisons embed a similar sensitivity to the state of the business cycle can we ensure that any observed differences are due to the extra information embedded in one measure rather than the other.

To assess this, start by taking the first derivative of the earlier job-filling rate equation to obtain:

$$\frac{d\frac{H}{V}}{d\theta} = -\alpha\mu\theta^{-(\alpha+1)}.$$

Next, recall that the measure of labor market tightness based on the standard measure of unemployment ($\theta_t^s = V_t/U_t^s$) is scaled differently from that based on the generalized measure of effective searchers ($\theta_t^g = V_t/U_t^g$). To be precise, the former has a mean of 0.58 and a standard deviation of 0.26, while the latter has a mean of 0.26 and a standard deviation of 0.71. So rather than comparing the responsiveness of the job-filling rate to a one-unit change in each measure, it would be more of an apples-to-apples comparison to evaluate how the job-filling rate responds to a one standard deviation change in each measure of labor market tightness, σ_θ. I call this the "cyclical sensitivity of the job-filling rate" and evaluate it at the mean level of tightness, $\bar{\theta}$, as follows:

$$\textit{Cyclical sensitivity of job-filling rate} = \sigma_\theta \times \left.\frac{d\frac{H}{V}}{d\theta}\right|_{\theta=\bar{\theta}} = -\sigma_\theta \alpha \bar{\theta}^{-(\alpha+1)}.$$

Table 2 shows how I calculate the cyclical sensitivity of the job-filling rate. The first four columns show the inputs into my calculations, and the final column shows the calculated sensitivity.

Importantly, notice that in panel A, the estimated sensitivity of the forecasts generated using the standard measure of unemployment is quite different from that calculated using the generalized measure of search activity. The point is that the relationship between job-filling, workers' search, and firm search used to compare these two measures is quite different in each case. And that in turn is the sense in which the comparisons calculated by the authors are not apples-to-apples comparisons.

By contrast, the panel B shifts the analysis to using the estimated matching function coefficients that I focus on in this comment. While the precise α's are quite different (that is α_s is different from α_g—reflecting differences in the scaling of the two competing measures of worker search, their implications), if instead we evaluate their consequences in terms of the effect of a one standard deviation change in search, the effects are quite similar. This suggests that the comparison based on the estimated coefficients is more of an apples-to-apples comparison. That comparison, in turn, found that the forecast of the job-filling rate based on the generalized measure of search effectiveness was almost identical to the forecast based on the standard measure of unemployment.

Finally, I should add that I'm not quite sure why the process by which the authors arrived at the α coefficients they impose are so different when

Table 2. Evaluation of the Cyclical Sensitivity of the Job-Filling Rate

	α *coefficient*	μ *coefficient*	*Average tightness* $\bar{\theta} = \overline{(V/U)}$	*Standard deviation of tightness* σ_θ	*Cyclical sensitivity of job-filling rate* $= -\sigma_\theta \alpha \bar{\theta}^{-(\alpha+1)}$
Panel A: Generating forecasts using imposed coefficients					
Using standard measure of unemployment	0.49	0.80	0.58	0.26	–0.23
Using generalized measure of search activity	0.57	0.48	0.26	0.71	–0.16
Panel B: Generating forecasts using estimated coefficients					
Using standard measure of unemployment	0.25	0.83	0.58	0.26	–0.10
Using generalized measure of search activity	0.44	0.53	0.26	0.71	–0.11

Source: Author's calculations.

using a measure of worker search based on the standard versus the generalized measure. The estimation is relatively opaque, and it is surely worth exploring why highly correlated series yield such different estimates of the cyclical sensitivity of the job-filling rate.

CONCLUSION This comment has dived pretty deep into the weeds, so it's worth panning back to the big picture. Abraham, Haltiwanger, and Rendell propose a new measure of labor market search that at a conceptual level has a lot to recommend it. The reality, however, is that it yields a measure that is remarkably highly correlated with a more standard measure like unemployment. The particular empirical exercise carried out in the paper asks which measure does a better job at predicting the job-filling rate. (It also asks which does a better job in matching the job-finding rate. For the sake of space, I've not dug into that measure here, but similar conceptual issues arise in analyzing competing predictions of that measure.)

Of course, one needs not just a measure of labor market slack but also a model if one is to predict the job-filling rate. And any model brings with it a set of auxiliary assumptions This comment has argued that these auxiliary assumptions are responsible for much of the difference in the forecasts generated by the two competing measures of labor market slack. My alternative set of tests are based on a different—and I would argue simpler, more transparent, more statistically coherent, and economically meaningful—set of auxiliary assumptions. And my alternative tests reveal that both measures of labor market slack yield almost identical forecasts of the job-filling rate. As such, there is little evidence that the new generalized measure of active searchers includes much useful information that's not already embedded in a more standard measure like unemployment. The new measure is neither meaningfully better nor worse than a more standard alternative.

GENERAL DISCUSSION Olivier Blanchard noted that the authors clearly made progress on the matching function but he was struck by the lack of discussion around the Phillips curve. He also commented that the analysis directly speaks to the paper in this volume by Del Negro, Lenza, Primiceri, and Tambalotti, which finds that wages are less responsive to measures of slack.[1] However, Abraham, Haltiwanger, and Rendell's paper suggests that slack is likely mismeasured. Blanchard continued by noting the importance of capturing the heterogeneity of workers in the matching function. It may also be that wage bargaining associated with these different groups of searchers is not the same, resulting in a more complex relationship between slack and wage setting.

Giorgio Primiceri commented that to understand if this new measure of slack performs better with the standard Phillips curve correlation, one needs to examine data from the 1960s to the 1980s. Primiceri continued by noting that Ayşegül Şahin's discussion showed that the movement of the authors' new measure of slack across the business cycle is likely very similar before and after the 1990s. So it is unlikely that this new measure of slack would solve the puzzle of the post-1990s Phillips curve correlation.

James Stock said he is unsure about the paper's argument that the Beveridge curve is more stable and not shifting up with this new measure of slack. While roughly vertical in the way it is graphed, it is not obvious whether it has less instability and appears to shift to the right during the relevant period.

Ricardo Reis wondered if the authors could relate their paper to the Stansbury and Summers paper in this volume, which finds a decline in worker bargaining power.[2] In a strict Diamond-Mortensen-Pissarides model,

1. Marco Del Negro, Michele Lenza, Giorgio E. Primiceri, and Andrea Tambalotti, "What's Up with the Phillips Curve?," *Brookings Papers on Economic Activity*, Spring (2020).

2. Anna Stansbury and Lawrence H. Summers, "The Declining Worker Power Hypothesis: An Explanation for the Recent Evolution of the American Economy," *Brookings Papers on Economic Activity*, Spring (2020).

one would think of the Beveridge curve separately from the bargaining of the surplus.[3] However, in a McCall model of the labor market where a change in the bargaining power of workers can impact their search decisions, one may find, as a result, a different relationship between the amount of vacancies and the amount of job seekers.[4]

Katharine Abraham and John Haltiwanger made several points in response to Justin Wolfer's discussion of the relative performance of forecasts based on generalized versus standard measures of labor market tightness. (Further details regarding the authors' response to these comments can be found in online appendix D.) Abraham and Haltiwanger observe, first, that Wolfers's analysis does not recognize the trade-off in how well the standard model tracks the job-filling rate versus how well it tracks the job-finding rate among the unemployed. Using Wolfers's ordinary least squares (OLS) estimates of the matching function elasticity, the standard model performs almost as well as the generalized model with regard to the job-filling rate. But, using those same elasticity estimates, the relative performance of the standard model for tracking the job-finding rate among the unemployed is significantly *worse*—applying Wolfers's OLS estimates to tracking the job-finding rate among the unemployed yields an RMSE for the standard model that is three times as large as that for the generalized model.

Another issue Abraham and Haltiwanger note with Wolfers's analysis is that it uses the same data to estimate model parameters and then evaluates their performance through the lens of a forecasting perspective. As discussed in more detail in online appendix D, when Wolfers's approach is modified so that models are fit with data through 2007 and then used to forecast outcomes in later years, the result is a generalized matching function that significantly outperforms the standard matching function both for predicting the job-filling rate and for predicting the job-finding rate among the unemployed. Finally, Abraham and Haltiwanger argue, the OLS approach Wolfers suggests for estimating the elasticity of the matching function yields estimates that are inherently biased. This is not true of the instrumental variables (IV) estimates developed in the paper.

3. Peter A. Diamond, "Wage Determination and Efficiency in Search Equilibrium," *Review of Economic Studies* 49, no. 2 (1982): 217–27; Dale T. Mortensen, "The Matching Process as a Noncooperative Bargaining Game," in *The Economics of Information and Uncertainty*, ed. J. J. McCall (Chicago: University of Chicago Press, 1982); Christopher A. Pissarides, "Short-Run Dynamics of Unemployment, Vacancies, and Real Wages," *American Economic Review* 75, no. 4 (1985): 676–90.

4. John Joseph McCall. "Economics of Information and Job Search," *Quarterly Journal of Economics* 84, no. 1 (1970): 113–26.

Using the IV estimates, the generalized model again significantly outperforms the standard model in tracking both the job-filling and the job-finding rates.

Haltiwanger addressed Şahin's and Blanchard's comments noting no incorporation of the Phillips curve in the analysis. Haltiwanger said that the paper produced an alternative measure to better capture slack in the labor market. Hopefully, there is consensus that the standard matching function is mis-specified. According to the data, a large portion of job seekers come out of the labor force and from among the employed, which is not captured by the standard matching function. Given this fact, their paper builds an alternative generalized matching function where the heterogeneity of job seekers is captured. Haltiwanger continued by saying that they are sympathetic to working out the details around their new measure and taking the next step to incorporate it into the Phillips curve framework. The new measure may have important implications for wage and price dynamics that should be explored in the future.

Haltiwanger responded to Reis's question by saying that if one were to pursue a Diamond-Mortensen-Pissarides model in the context of the Phillips curve, one would need to take a generalized approach like they do in the paper.

Haltiwanger continued by addressing Stock's question and part of Şahin's comment around normalizing measures. The authors normalized their measures because they think that the intercept of the hiring function should be different across specifications. The standard model should not have the same intercept as the generalized model. The authors are interested in removing those intercept differences, so they normalized their measures.

Haltiwanger then turned to another part of Şahin's comment around the number of subgroups needed to calculate a generalized matching function. He acknowledged that it's not clear whether one needs twenty-two groups for the sample period covered by the paper. The most important subgroups in this sample period are decomposing the unemployed into short-term and long-term unemployed and including and decomposing the group of those out of the labor force into want a job and other. However, over a longer sample period (e.g., including the entire post-WWII period), it is likely that additional subcategories matter. In considering this issue, it is noteworthy that there are large differences in relative search intensities across the detailed groups but for this to matter there also need to be changes in composition across the groups over time.

MIKKEL PLAGBORG-MØLLER
Princeton University

GIOVANNI RICCO
University of Warwick

LUCREZIA REICHLIN
London Business School

THOMAS HASENZAGL
University of Minnesota

When Is Growth at Risk?

ABSTRACT This paper empirically evaluates the potentially nonlinear nexus between financial indicators and the distribution of future GDP growth, using a rich set of macroeconomic and financial variables covering thirteen advanced economies. We evaluate the out-of-sample forecast performance of financial variables for GDP growth, including a fully real-time exercise based on a flexible nonparametric model. We also use a parametric model to estimate the moments of the time-varying distribution of GDP and evaluate their in-sample estimation uncertainty. Our overall conclusion is pessimistic: moments other than the conditional mean are poorly estimated, and no predictors we consider provide robust and precise advance warnings of tail risks or indeed about any features of the GDP growth distribution other than the mean. In particular, financial variables contribute little to such distributional forecasts, beyond the information contained in real indicators.

Conflict of Interest Disclosure: Mikkel Plagborg-Møller is an assistant professor of economics at Princeton University; Lucrezia Reichlin is a professor of economics at the London Business School, chair and cofounder of Now-Casting, an advanced statistical model to monitor macro conditions, a nonexecutive director for the U.K.-based auto and home insurance company Ageas, a nonexecutive director for Eurobank (Greece), and a nonexecutive director for the Italian publisher and holding company Messaggerie Italiane; Giovanni Ricco is an assistant professor of economics at the University of Warwick, an associate researcher at OFCE-Sciences Po (Observatoire français des conjonctures économiques, an independently and publicly funded research center), and a research affiliate at the Center for Economic and Policy Research; Thomas Hasenzagl is a Ph.D. student in economics at the University of Minnesota. Beyond these affiliations, and a grant from the National Science Foundation, the authors did not receive financial support from any firm or person for this paper or from any firm or person with a financial or political interest in this paper. They are currently not officers, directors, or board members of any organization with an interest in this paper. No outside party had the right to review this paper before circulation. The views expressed in this paper are those of the authors, and do not necessarily reflect those of the University of Minnesota, the London Business School, Now-Casting, Ageas, Eurobank, Messaggerie Italiane, the University of Warwick, Sciences Po, or Princeton University.

Following the Great Recession, there has been an increasing interest in understanding the relationship between financial fragility and the business cycle. Having failed to predict the crash, the economics profession has been trying to understand what was missing in standard macroeconomic models and what are the key indicators of stress in financial markets which may help forecast crises and identify the build-up of macroeconomic risks ahead of time. The research agenda involves not only prediction but also a revisitation of the earlier literature on financial frictions and the business cycle, pioneered by Bernanke and Gertler (1989), Kiyotaki and Moore (1997), and Bernanke, Gertler, and Gilchrist (1999), on the basis of the experience of the 2008 Great Recession.

This research goes beyond academia since it is potentially informative for macroprudential policy, which indeed focuses on the interaction between financial institutions, markets, and the wider economy. Such policies need to be grounded in theoretical and empirical knowledge of what are the appropriate tools for strengthening the resilience of the financial system to macroeconomic shocks and vice versa. Early warnings of growth fragility would allow monetary and fiscal policymakers to respond proactively to budding crises.

The structural literature has focused on two alternative classes of variables: those capturing the effect of an external financial premium (in line with models based on the financial accelerator) and those capturing balance sheet constraints such as household or bank credit, reflecting the idea that leverage is a main indicator of the accumulation of financial instabilities (Gertler and Gilchrist 2018). Price variables such as credit spreads are typically used as proxies for the external financial premium. In fact, there is some consensus that measures derived from different types of interest rate spreads can have predictive power for future economic conditions. For the United States, for example, the influential work of Gilchrist and Zakrajšek (2012) has proposed a measure of an excess bond premium that has been widely adopted in both academic and policy work.

A different but related line of research, pioneered by the Bank for International Settlements (BIS), has stressed the importance of the leverage cycle as an indicator of risk and used excess private credit as a measure of macrofinancial imbalances (Basel Committee for Banking Supervision 2010). Some studies have pointed at a correlation of excess growth in leverage and financial crises (Jordà, Schularick, and Taylor 2011, 2013; Schularick and Taylor 2012) and found that recessions preceded by financial crises are deeper and followed by slower recoveries (Reinhart and Rogoff 2009; Valencia and Laeven 2012).[1] However, this literature is mainly concerned with long-term features of the nexus between finance and the macroeconomy and on financial crises rather than recessions. At business cycle frequency, growth rates of credit aggregates are found to be procyclical and lagging (Giannone, Lenza, and Reichlin 2019). In a recent paper, Brunnermeier and others (2019) have pointed out that credit moves passively with output but that the negative correlation between credit spreads and output is mostly explained by the endogenous response of monetary policy.

Although the literature is very rich, few robust results have emerged from empirical studies about the extent to which financial variables can be used to predict economic activity. This confirms the conclusions of earlier work (Stock and Watson 2003; Forni and others 2003; Hatzius and others 2010). In particular, three features of financial variables provide challenges to probing both the predictive and the causal relationships connecting them to the real variables. First, movements in financial variables are largely endogenous to the business cycle. Second, the dynamics of financial variables—and spreads in particular—are potentially nonlinear and may be related to the higher moments of the GDP distribution rather than just the central tendency. Finally, there is a great degree of heterogeneity among financial indicators. Different types of financial variables capture different mechanisms through which financial markets and the macroeconomy interact.

The idea that financial and economic conditions may be correlated nonlinearly has recently inspired a line of research which uses nonparametric methods in order to study the predictive distribution of GDP and its evolution in relation to financial conditions. Giglio, Kelly, and Pruitt (2016) and Adrian, Boyarchenko, and Giannone (2019a) estimate the predictive GDP distribution conditional on a synthetic index of financial conditions. This index aggregates variables capturing financial risk, leverage, and credit quality. For the United States, such an index is constructed by the Federal Reserve Bank of Chicago—the National Financial Conditions

1. A related but different line of research has identified a financial cycle with different characteristics than the business cycle but leading it and found that financial cycle booms either end up in crises or weaken growth (Borio and Lowe 2002; Drehmann, Borio, and Tsatsaronis 2012; Claessens, Kose, and Terrones 2012).

Index (NFCI). Both papers, focusing on US data, found that the lower quantiles of GDP growth vary with financial conditions while the upper quantiles are stable over time, therefore pointing to an asymmetric and nonlinear relationship between financial and real variables. New research is building on these ideas. Recent contributions are by Kiley (2018), Adrian, Boyarchenko, and Giannone (2019b), Loria, Matthes, and Zhang (2019), Brownlees and Souza (2019), and Figueres and Jarociński (2020).

As proposed by Adrian and others (2018), the evaluation of the predictive GDP distribution can be used to define the concept of *growth at risk*, defined as the value of GDP growth at the lower fifth percentile of the predicted growth distribution, conditional on an index of financial stress. This concept has been adopted by policy institutions in many different countries to monitor risks.[2] The appeal of this approach to policy work, in particular macroprudential, is that it provides a framework in which forecasting can be thought of as a risk-managing exercise.[3]

The value of this framework for policy in practice rests on whether the dynamics of the moments of the conditional distribution of GDP can be captured with some degree of precision and on whether there is some out-of-sample predictability for moments other than the mean. In a recent paper, Reichlin, Ricco, and Hasenzagl (2020) evaluate the out-of-sample performance of an aggregate indicator of financial stress and of some key financial variables for the GDP distribution, using the nonparametric approach of Adrian, Boyarchenko, and Giannone (2019a), and found little evidence of predictability beyond what can be achieved using timely indicators of the real economy. In this paper we broaden this analysis in several directions by asking three questions.

First, we want to assess the marginal role of financial variables in estimating and predicting the conditional distribution of GDP once we condition appropriately on available monthly macroeconomic information. Our conjecture is that monthly macroeconomic and financial variables co-move strongly at the contemporaneous level and that a large part of what is revealed by the NFCI reflects some joint information. This of course would not be the case if financial markets primarily reflected forward-looking information, a feature which cannot be assumed and must be tested.

2. See, for example, Prasad and others (2019) for a description of the use of this method at the IMF.

3. See Greenspan (2004) and Kilian and Manganelli (2008).

Second, we want to evaluate whether nonlinearities in the predictive distribution can be effectively exploited for forecasting and whether the dynamics of moments other than the mean can be precisely estimated. We believe that both evaluations are important for understanding whether the growth-at-risk framework can be used in practice for macroprudential policy. The out-of-sample evaluation takes into consideration overall uncertainty: stochastic, estimation, and model uncertainty. Parameter uncertainty—that is, uncertainty conditional on a particular assumed model—can be evaluated in-sample. For the first purpose we use the nonparametric method proposed by Giglio, Kelly, and Pruitt (2016) and Adrian, Boyarchenko, and Giannone (2019a), while for the second purpose we use a fully parametric implementation of their approach. The motivation for using two different models is that the nonparametric approach very flexibly captures nonlinearities without relying on particular functional forms, but, unlike the parametric method, it cannot easily be used to assess the statistical uncertainty surrounding the estimation of the moments of the growth distribution. We view the two approaches as complementary.

Third, we assess the potentially different roles of individual financial variables in estimating the moments of the conditional distribution by considering a variable selection algorithm. The motivation here is that—as has been observed by Reichlin, Ricco, and Hazenzagl (2020)—financial variables have very different dynamic properties so that, by aggregating predictors into financial and real indexes as done in the literature, some information can be lost. An approach that allows individual variables to enter the model in a flexible way may therefore be of interest. Moreover, understanding which specific economic variables carry information about the distribution of GDP growth would allow policymakers and academics to hone in on specific mechanisms of growth fragility. We consider both US data and a panel of twelve other Organisation for Economic Cooperation and Development (OECD) countries. This allows us to consider more than a few recessionary events in our sample. For the United States, for which we have a richer data set, we perform the analysis both separately and in combination with other countries' data.

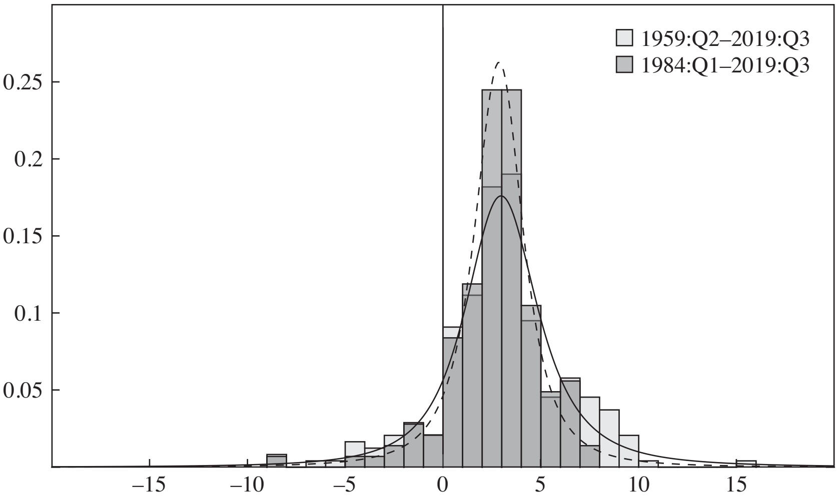
The overall conclusion of our analysis is pessimistic on the ability of the data to tell us something more than the evolution of the conditional mean. All other time-varying moments are imprecisely estimated. Moreover, both the out-of-sample analysis and the in-sample results point to very little additional predictive power of financial variables for other moments and for all moments at longer horizons. This remains true in a real-time nowcasting exercise where we take into account the timeliness of financial variables relative to other data, since survey data are almost as timely and highly correlated with macroeconomic data. Finally, when single variables are allowed to enter into the model flexibly, these results are confirmed for both credit spreads (prices) and credit aggregates (quantities), although our methods cannot rule out that some interaction between spreads and credit is at work.

In section V, we run the real-time experiment over the recent COVID-19 lockdown episode in the first months of 2020. In this case, the model with financial variables does provide a more timely indication of the directional movement of the GDP growth distribution, relative to models that only condition on nonfinancial data. However, no model gets close to accurately predicting the severe magnitude of the downturn. Moreover, the COVID-19 episode has no bearing on the question of whether financial variables are helpful predictors outside very short forecast horizons.

At a more general level, our analysis confirms the older literature's results of the lack of predictive power of financial variables for the real economy, but we show that this finding carries over to an approach that in principle is capable of capturing nonlinearities and tail risks. Our findings suggest that markets do not anticipate the timing of the recession and they price the risk only once they see it. In other words, the onset of a recession comes as a surprise to seemingly all agents in the economy. This blindness can be interpreted as revealing that information is rapidly available to all, but rare events such as recessions are fundamentally unforecastable. Importantly, our results do not imply that macroprudential policy should give up on limiting the accumulation of financial fragilities, since it is likely that those fragilities amplify the damage to the real economy once recessions do occur. However, this is not a question that we can evaluate using the methods in this paper.

The sections of the paper are organized around the questions we ask. After presenting some motivating facts in section I, section II asks the question of whether financial variables have specific forward-looking information that can inform an out-of-sample predictive relationship with the mean or higher moments of the GDP distribution. We also assess whether financial variables have predictive power for the GDP distribution during the nowcasting period, where we consider their timeliness advantage with respect to real economic indicators. Section III asks how precisely the moments of the predictive distribution of GDP growth, conditional on real and financial factors, can be estimated in-sample. As in section II, we use as predictors both a global factor that includes joint real and financial information and a financial factor that includes the financial information

Figure 1. Annual Real GDP Growth



Sources: FRED-QD and authors' calculations.
Notes: Histograms of annual real GDP growth over the samples 1959:Q2–2019:Q3 and 1984:Q1–2019:Q3. The fitted distributions are computed by adopting the flexible skew *t*-distribution developed by Azzalini and Capitanio (2003).

orthogonal to the global factor. Section IV abandons the factor-based predictors and instead asks whether there are any specific individual economic variables that are able to explain the dynamics of GDP growth moments. As a case study, we evaluate the nowcast of the GDP growth distribution in the recent COVID-19 lockdown episode in section V. Section VI concludes.

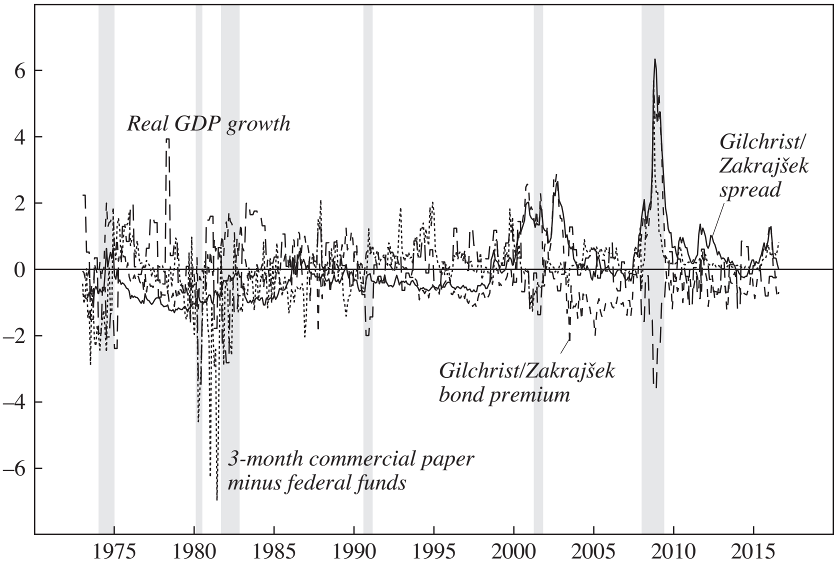
I. A Few Motivational Facts

In this section we present a few facts that motivate the analysis of the paper.

I.A. Fact 1: Economic Fluctuations Are Asymmetric over the Business Cycle

Figure 1 shows that the distribution of US GDP growth exhibits some skewness and fat tails. The figure plots the histograms of annual real GDP growth over the samples 1959:Q2–2019:Q3 and 1984:Q1–2019:Q3 and the associated fitted distributions. The dark area marks the overlapping segments. Growth in both subsamples exhibits skewness and heavy tails, although arguably to varying degrees. Indeed the literature has suggested that recessions can be described as a combination of a negative first-moment

Figure 2. Financial Stress Indicators and GDP Growth Rates



Sources: FRED-MD, FRED-QD, and Gilchrist and Zakrajšek (2012).

(mean) shock and a positive second-moment (uncertainty) shock (Bloom 2014) or as negative third-moment (skewness) shocks (Bloom, Guvenen, and Salgado 2019), and fat tails have been found to be a feature of GDP distribution in many advanced economies (Fagiolo, Napoletano, and Roventini 2008).

This fact motivates an analysis which is based on estimation and forecasting of moments other than the mean of the predictive GDP distribution.

I.B. Fact 2: Financial Condition Indicators and Spreads Are Highly Negatively Correlated with Output Growth at the Time of Recessions

Figure 2 shows a clear negative correlation between spreads and GDP growth around recessions (although the relation is unstable over the sample). The figure plots quarterly annualized GDP growth for the period from 1973:Q1 to 2016:Q3 against three credit spreads that have been considered in the literature as measures of financial risk (Gilchrist and Zakrajšek 2012).

This chart suggests that the asymmetry in the business cycle for output growth is associated with the asymmetry in the behavior of credit spreads.

The latter increase sharply in coincidence or just prior to an economic contraction, while there is no symmetric movement in these variables during booms. The intriguing suggestion is that, by conditioning on these variables, it would be possible to capture higher moments of the GDP conditional distribution. As discussed above, this idea has been the inspiration for the literature that has explored the predictive power of financial variables for moments other than the mean, which we seek to evaluate in this paper.

I.C. *Fact 3: Movements in Financial Conditions Are Largely Endogenous and Related to Output Growth*

Financial time series and macroeconomic variables share a pronounced contemporaneous common component. Figure 3 reports the quarterly average of the monthly NFCI and of a business cycle index computed from a large set of monthly macroeconomic indicators.[4]

The two synthetic aggregate indicators of financial and macroeconomic variables exhibit a very clear pattern of co-movement. The strong correlation emerging from the plot indicates that movements in financial indicators are possibly endogenous and contemporaneous to business cycle fluctuations.

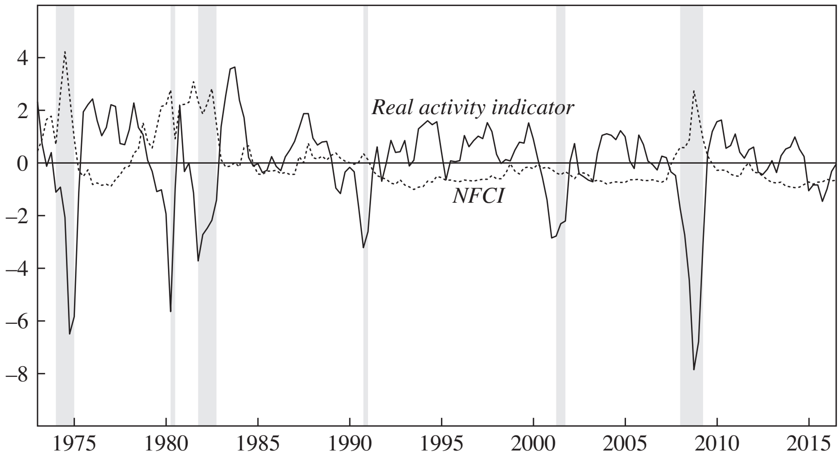
This fact suggests that, in order to establish the role of financial variables for predicting the GDP distribution, one should control for the common and contemporaneous component (what we define as the "global factor") and focus on the additional "marginal" information available in the financial indicators (the "financial factor"). This is what our analysis will do.

I.D. *Fact 4: Different Types of Financial Variables Have Heterogeneous Dynamics along the Business Cycle*

Figure 4 provides a more disaggregated view of financial stress by plotting the NFCI and its components. The chart suggests that the NFCI

4. The business cycle index is computed as the first common factor to all of the variables in the FRED-MD data set, except the ones classified as financial. Online appendixes S.A and S.B provide details on the estimation of the factor. The NFCI index is a synthetic indicator computed as a common factor extracted from 105 mixed-frequency—weekly, monthly, and quarterly—financial variables. It averages four categories of data: credit quality, risk, nonfinancial, and financial leverage. All variables are transformed to stationarity and standardized. For a description of the NFCI (variables considered and methodology), see Brave and Butters (2012) and the Federal Reserve Bank of Chicago's dedicated website: https://www.chicagofed.org/publications/nfci/index. Both factors are estimated by maximum likelihood following Doz, Giannone, and Reichlin (2012) and averaged across quarters. Table S.4 in the online appendix S.D reports the full set of estimated values for the model coefficients.

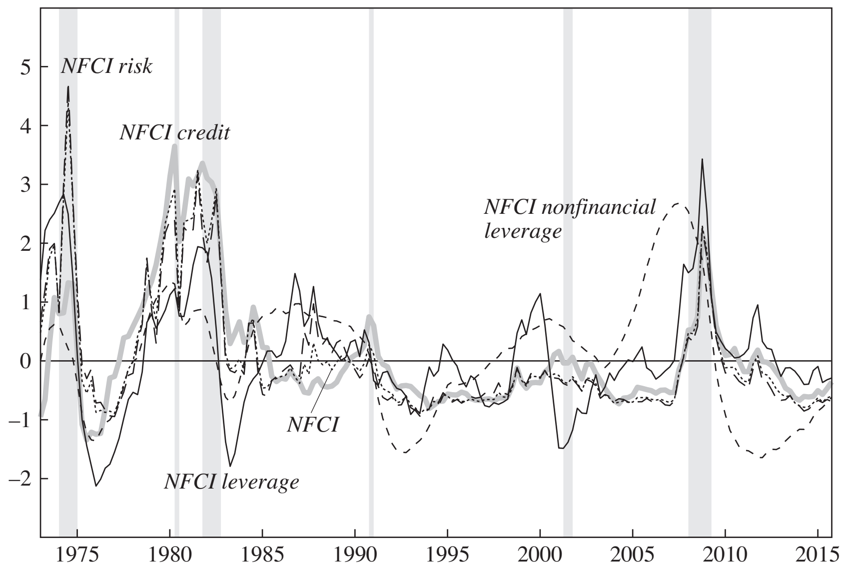
Figure 3. Business Cycle and Financial Condition Indexes



Source: Authors' calculations.

Note: The chart plots an index of real activity extracted as a common factor from a large set of macroeconomic variables and excluding financial variables against the NFCI. The time sample is 1973:Q1–2016:Q3.

Figure 4. Heterogeneous Dynamics of Financial Indicators



Source: FRED-QD.

aggregates components with heterogeneous dynamic characteristics, potentially reflecting different forms of fragility in the financial system. It shows that the aggregate NFCI dynamics reflect mainly the risk and credit components, while nonfinancial leverage follows a smoother cyclical pattern, and financial leverage exhibits some higher-frequency idiosyncratic dynamics.

Indeed, different indicators of stress capture different aspects of financial frictions, which may be relevant at different moments in time—either preceding, contemporaneous to, or following the financial crisis.[5]

This fact motivates our analysis of the role of individual variables in predicting the moments of the conditional distribution of GDP growth.

II. Predicting Growth at Risk

In this section we assess whether financial variables aggregate forward-looking information that helps predict the distribution of future GDP growth. In particular, we are interested in teasing out information about the future path of output and its moments in excess of the contemporaneous information provided by other macroeconomic indicators. Toward this aim, we consider the marginal gain in the predictive distributions for GDP growth (and its moments) when financial-specific information is incorporated, relative to baseline models that only condition on the global common component in real and financial data.

We provide both an out-of-sample exercise—forecasting one quarter and four quarters ahead—and a fully real-time monitoring of risks to GDP growth with a realistic data release calendar, encompassing macroeconomic and financial variables. It is worth observing that the out-of-sample exercise provides an overall summary of the performance of the model by factoring in several types of uncertainty, excluding the uncertainty about data itself that is a component of the flow of revised data releases. The real-time exercise takes the latter dimension of uncertainty partially into account since it is based on a realistic calendar of data releases mimicking the information flow.

The results are overall negative. The inclusion of financial-specific information does not improve the mean squared forecast error of the model, nor does it help capture the dynamics of any of its moments. However, financial variables appear (very marginally) to help in pinning down the common contemporaneous information in real time.

5. See Bernanke (2018) for an analysis of the 2008 recession in the United States.

II.A. The Evolution of Out-of-Sample Growth Movements

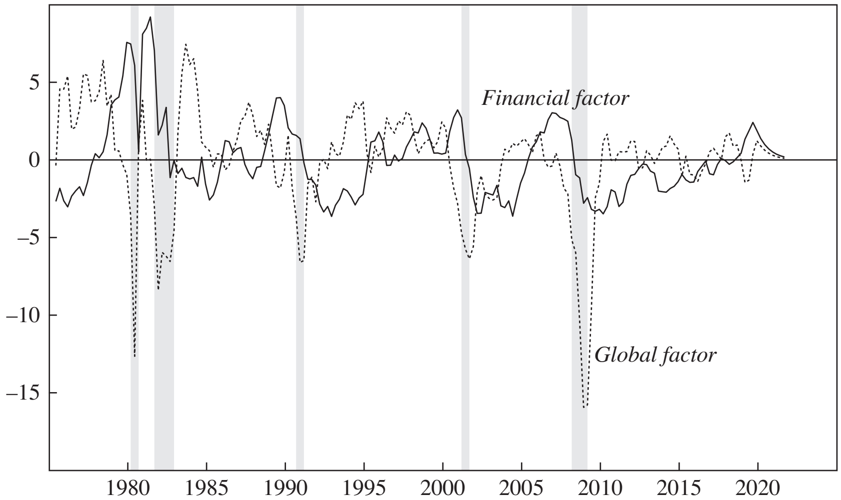
We first ask the following questions: How do the moments of the predictive distribution vary over time? Do financial variables capture shifts in the predictive mean, variance, or higher moments of the GDP distribution? Is it possible to predict an increase in GDP growth vulnerability out of sample? This exercise focuses on short-to-medium horizons and tries to gauge the overall abilities of the models in assessing risks to GDP growth. Importantly, while providing an assessment of the models' performance against the several sources of uncertainty—stochastic, estimation, and model uncertainty—it abstracts from the data uncertainty that characterizes data releases in real time. We integrate this last source of uncertainty in the subsequent real-time exercise.

DATA AND MODEL The first step in our exercise is the estimation of common factors from a large panel of variables. Specifically, we extract two indexes of commonalities. The first factor, which we refer to as the *global factor*, is common to all the variables in the McCracken and Ng (2016) Federal Reserve Economic Data Monthly Database (FRED-MD) data set, including real, financial, monetary, and price variables. The second factor, which we refer to as the *financial factor*, is only common to the financial variables and is by definition orthogonal to the global factor. Figure 5 plots the two factors over the sample period. Online appendix S.A provides details on the factor models adopted to estimate the factors.[6] Table S.1 in online appendix S.B provides details on the data set and on the assumptions adopted to estimate the factors.

The key difference from the analysis of Adrian, Boyarchenko, and Giannone (2019a) is that, while they adopt the NFCI as the main indicator of financial conditions, we separate the information contained in the global factor and the orthogonal financial factor. Reichlin, Ricco, and Hasenzagl (2020) observe that the NFCI is largely endogenous to economic conditions in the United States and that it has high correlation with a factor extracted from nonfinancial variables only (as also shown in figure 3). This observation motivates our choice to adopt a global indicator of economic conditions as well as a financial-specific factor that could, in principle, capture independent forward-looking information about the moments of the predictive distribution of GDP growth that is not obtainable from current economic conditions.

6. Figure S.1 in online appendix S.C reports the estimated loadings for the factor model with a global factor and a financial factor.

Figure 5. Global and Financial Factors



Source: Authors' calculations.
Note: The time sample is 1975:Q2–2021:Q3. The values between 1975:Q2 and 2019:Q3 are in-sample estimates of the factors and the values between 2019:Q4 and 2021:Q3 are out-of-sample forecasts.

We employ the factors as predictors in the nonparametric quantile regression framework of Adrian, Boyarchenko, and Giannone (2019a). To compare the predictive content of the two factors, we consider three empirical specifications. We model annualized cumulative GDP growth at the one-quarter-ahead and four-quarter-ahead horizons as being driven by, respectively,

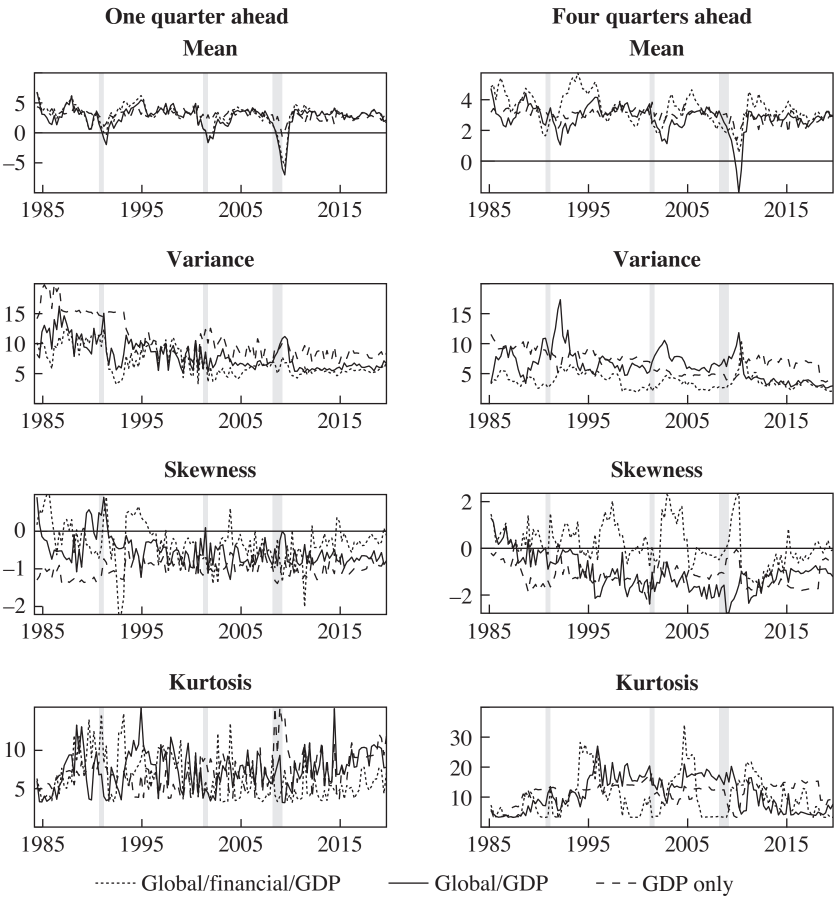
(model 1) GDP growth at time t;

(model 2) GDP growth at time t and the economic activity global factor at time t; and

(model 3) GDP growth at time t and both the global and the financial factors at time t.

We first estimate the factor model using data from 1975:Q2 to 1984:Q1. We then iteratively estimate the predictive distributions of GDP growth one and four quarters ahead, expanding the estimation sample, one quarter at a time, until the end of the sample in 2019:Q3. In every quarter of the out-of-sample period, we apply the nonparametric prediction approach of Adrian, Boyarchenko, and Giannone (2019a). This involves first estimating the relationship between the percentiles of future GDP growth and the predictors using quantile regressions. Then we smooth out the predictive

Figure 6. Out-of-Sample Forecasts: Time Evolution of the Predictive Distribution of GDP Growth



Source: Authors' calculations.

Note: Time evolution of the four moments of the one-quarter-ahead predictive distribution of GDP growth, from 1984:Q1 to 2019:Q3, for three models: including the global factor, financial factor, and GDP; including the global factor and GDP; and GDP only.

distribution by fitting a flexible family of distributions to the estimated conditional percentiles, allowing for both skewness and heavy tails. The details of the prediction procedure are described in online appendix S.A.

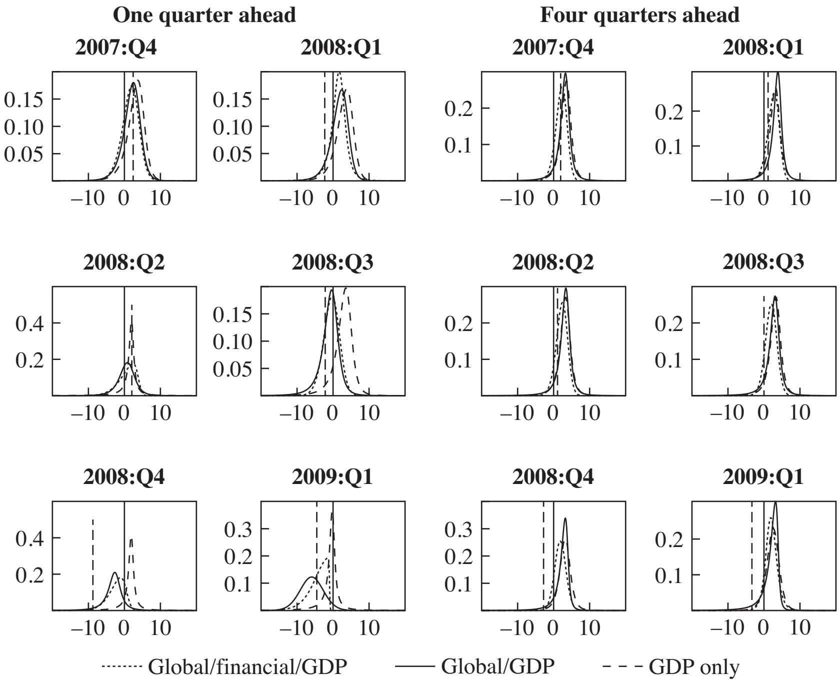
RESULTS Regardless of the predictors used, the models fail to provide noticeable advance out-of-sample signals of the likelihood or severity of recessions. Figure 6 shows the first four moments of the forecast distribution of GDP growth at horizons $h = 1$ and $h = 4$. By breaking down the predictive distribution into different moments, we aim to show what features of the distribution of GDP growth are predictable, if any. The figure compares three models: one that includes the global factor, the financial factor, and GDP; one that includes the global factor and GDP; and one that includes lagged GDP only.

At the one-quarter-ahead horizon ($h = 1$) shown on the left, the distributions of both models that incorporate factors show a sharp decrease in the mean around the period of the Great Recession, but importantly, the model incorporating the financial factor does not seem to have an informational advantage. Strangely, the model not incorporating the financial indicator seems to capture an increase in the variance related to the Great Recession, albeit with some delay. In fact, the movement in the variance lags the 2008 recession by a few quarters, and it results from the incorporation into the model, with a quarter of delay, of the spike in spreads in the fourth quarter of 2008. Also, the increase is not remarkable when compared to the level of the forecast variance in the 1990s. Skewness and kurtosis apparently move over the sample but with patterns that are not easy to interpret or to relate to economic contractions.

At the four-quarter-ahead horizon ($h = 4$) shown on the right, the findings are in line with those discussed for $h = 1$ but the reactions to contractions are even more delayed. Interestingly, only the model with the global factor forecasts substantial contractions in GDP at the four-quarter horizon around recessionary periods, although with long delay. Higher moments do not exhibit interpretable patterns. This raises doubts about the ability of the models to correctly capture the dynamics of these moments, at least out-of-sample, an issue we will return to in section III.

We now zoom in on the Great Recession period. Figure 7 reports the two predictive distributions at different points in time (2007:Q4–2009:Q1), for $h = 1$ and $h = 4$, before and during the Great Recession for the three different models. None of the models seem to predict the crisis. At horizon $h = 1$ (the set of graphs on the left), all the models fail to capture the onset of the economic downturn in 2008:Q1, and they all assign a low probability to it. As financial stress spikes up in the fourth quarter of 2008, the conditional forecast of both models that include the global factor fans out, attaching higher likelihood to a wider range of events. At horizon $h = 4$ (the set of graphs on the right), all models seem to do equally poorly in capturing the shift in economic conditions. Although the model that only conditions on lagged GDP performs particularly poorly, the two models

Figure 7. Out-of-Sample Forecasts: Predictive Distributions during the Great Recession



Source: Authors' calculations.

Note: Quarter-by-quarter evolution of the predictive distributions in the period of the Great Recession, from 2007:Q4 to 2009:Q1, for three models: including the global factor, financial factor, and GDP; including the global factor and GDP; and GDP only. The charts report also the realization of annualized GDP growth one and (cumulatively) four quarters ahead, respectively.

incorporating factors yield very similar predictive distributions. Indeed, the model that also incorporates financial variables seems to have little informational advantage.

A more systematic evaluation of the distributional forecast accuracy by analyzing the models' predictive scores confirms the minuscule predictive content of the financial factor. This is shown in online appendix S.D in figure S.2. The predictive score is high if a model attaches a high likelihood to the value of GDP growth that is actually realized (see the formal definition in online appendix S.A). While at $h = 1$ the two models have nearly indistinguishable predictive scores, at $h = 4$, the model incorporating the financial factor seems to have a very small advantage over the model with the global factor only. Yet its performance does not uniformly dominate the second model over the sample.

SUMMARY An explorative out-of-sample analysis indicates that financial variables help only very marginally in improving the performance of a model that already includes a real activity indicator, computed as the common factor of a large panel of real macroeconomic variables. Interestingly, the movements in higher moments of the forecast seem not to be very informative.[7] In particular, skewness and kurtosis do not show any interpretable movement around recessions. This suggests that growth vulnerability is a story about the mean and possibly volatility of growth, rather than about time variation in the probability of extreme events. We return to this issue in section III, where we will be able to characterize the statistical uncertainty associated with the estimation of each time-varying moment. In the next subsection we explore the specific informational content of financial indicators and their relations with real variables, their timeliness, and the heterogeneity across financial variables.

II.B. Real-Time Monitoring of Risks to Growth

To assess the predictive ability of the quantile regression model in real time, we turn to nowcasting, that is, predicting the current quarter value of GDP growth ($h = 0$). We will also continue to consider the one-quarter-ahead forecast horizon ($h = 1$). Although these horizons are too short-term for the practical implementation of macroprudential policies, they are relevant for prediction since the literature has shown that, generally, there is very little predictability for the mean of GDP growth beyond one quarter (Giannone, Reichlin, and Small 2008). Additionally, monetary and fiscal policy may be able to respond within the quarter in some cases. Finally, our results so far seem to indicate that the model has limited predictive ability at longer horizons anyway.

DATA AND MODEL In this exercise we update the factors and hence the forecast and nowcast in relation to a calendar of data releases, in the tradition of the nowcasting literature. First, we construct a set of real-time data vintages from the Archival Federal Reserve Economic Data (ALFRED) database. The data series that we include were chosen to closely resemble the FRED-MD data set, given data availability constraints of the real-time data. The real-time vintages for some variables only become available after the beginning of the forecasting exercise. Those variables are added to the exercise once they become available. As we did above, we extract a number of common factors from those vintages. Beyond the global factor

7. This is consistent with the findings of Adrian, Boyarchenko, and Giannone (2019a).

(common to all the variables) and the financial factor (common to the financial variables only and orthogonal to the global factor), we also consider a nonfinancial factor, computed from the subset of the data set that excludes financial variables.

The calendar of data releases uses the average release lag for each variable. In the out-of-sample exercise, we then iterate over the release calendar, position ourselves at each release date, and perform the following two-step procedure:

(step 1) We estimate the factors using an expectation-maximization (EM) algorithm. Then we average the monthly factors to get quarterly factors.

(step 2) We apply the nonparametric forecast approach of the previous subsection to quarterly data up to the current quarter. Using this approach, we construct predictive distributions for current quarter and next quarter GDP growth.

We consider the following three sets of predictor variables:

(model 1) global factor only;

(model 2) global factor and financial factor; and

(model 3) nonfinancial factor only.

We construct quarterly versions of the factors as averages of the factors estimated in a monthly nowcasting model (Giannone, Reichlin, and Small 2008). We begin the out-of-sample forecasting exercise in 2005:Q1. For each data release we estimate the factors and the quantile regression parameters using an expanding data set starting in 1980:Q1.

Some of the financial variables included in our real-time exercise—stock indexes, oil price, exchange rates, interest rates, and spreads—are available at daily or higher frequency. However, they enter the model only as end-of-the-month values on the first day of the following month. This, while being a blunt approximation of the information flow, still affords these financial variables an informational advantage by including them in the model before any real and nominal variable, for the month of interest. Table 1 shows the average lag of the release of the most important groups of variables that we use in the exercise. Table S.1 in online appendix S.B shows all the variables included in the data set, their average release lag, and the factors on which they load. By employing the growth-at-risk framework, our methodology also allows for financial variables to affect higher moments of the GDP forecast, which could be particularly important in determining tail risks.

Comparing the short-term forecasting performance of a model that contains only the global factor and a model that contains both the global

Table 1. Groups of Variables Used in the Nowcast Exercise and Their Release Lags

Variable group	*Release lag*
Stock indexes, exchange rates, interest rates, and spreads	1
Institute for Supply Management indexes	1
Employment and earnings	5
Monetary aggregates	15
Industrial production and subcomponents	16
CPI, producer price index, and subcomponents	16
Housing starts, housing permits, and subcomponents	18
Personal consumption expenditure and real personal income	30

Sources: Archival Federal Reserve Economic Data (ALFRED) and authors' calculations.
Note: The lag variable is the approximate number of days between the last day of the reference month and the date at which the variable becomes available.

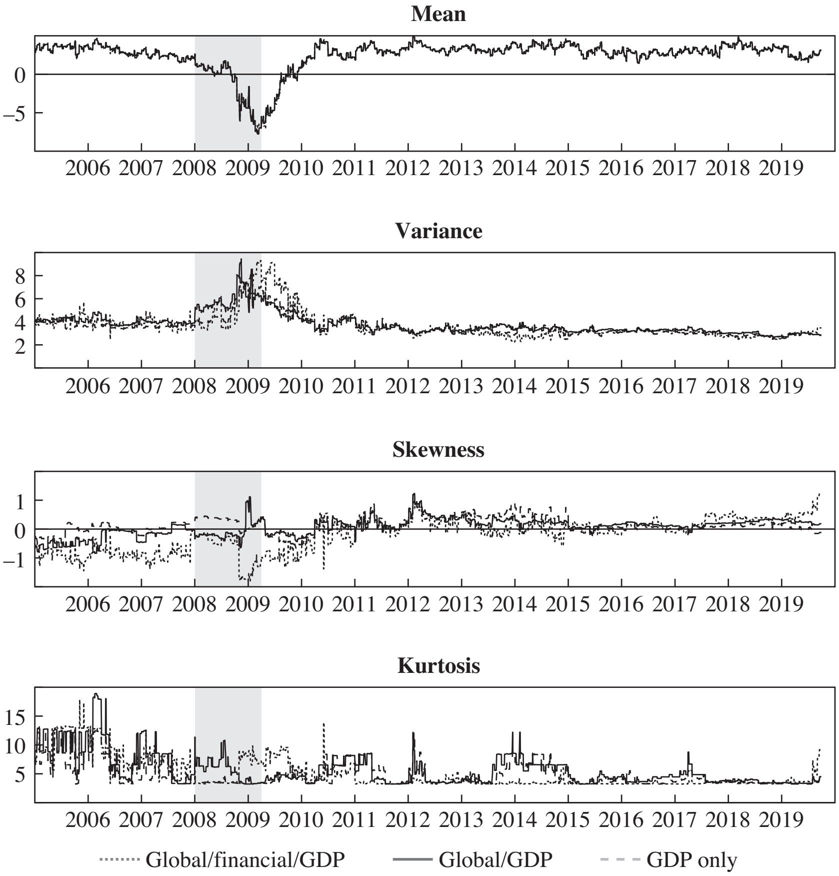
and financial factors allows us to study the additional information content of financial variables over and above what is common to all the other economic variables. Additionally, comparing the short-term forecasting performance of the model that contains only the nonfinancial factor helps assess the effects of financial variables on imputing the global factor.

RESULTS Financial variables help only very marginally for nowcasting, and only because they help to estimate the global factor more precisely. Figure 8 reports the evolution over time of the four moments of the predictive growth distribution at horizon $h = 0$. The top panel shows that the conditional means of the predictive distributions in all models are nearly identical. The global factor captures the co-movement between all variables, including the financial variables, and adding the orthogonal financial factor does not have a substantial effect on the mean of the predictive distribution. The model with the factor estimated using only nonfinancial variables provides a forecast for the mean that is nearly identical to that of the other models.

The models disagree more about the variance, skewness, and kurtosis of the predictive distributions. For example, in the middle of the Great Recession, the model with the financial factor shows an increase in kurtosis in 2008 and a spike in skewness early in 2009. While these features are not prominent in the sample, they may be an indication that the real-time model that incorporates financial variables captures some downside risks to growth, although with a delay.

Figure 9 shows that the early availability of financial variables does not translate into more accurate forecasts of the mean of the GDP distribution at short horizons. The top chart reports the root-mean-square forecast error of the three models, which depends only on the mean of the predictive

Figure 8. Nowcast of the Moments of GDP Growth

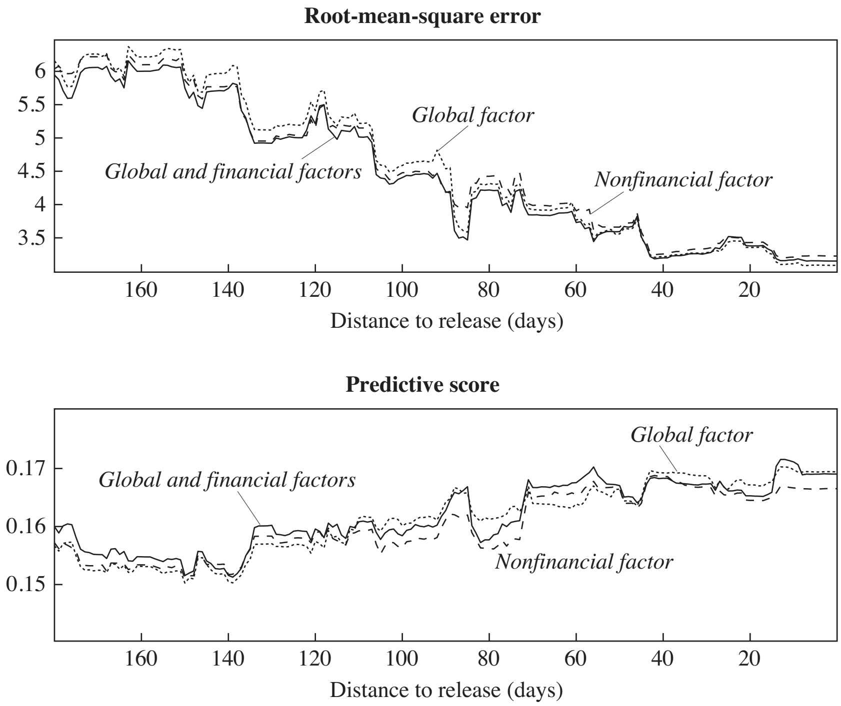


Source: Authors' calculations.

Note: Time evolution of the four moments of nowcast predictive distribution of GDP growth at $h = 0$ of quantile regressions with the global factor only, with the global and financial factors, and with the factor estimated using only nonfinancial variables, from 2005:Q1 to 2019:Q3.

distributions, as a function of the remaining time until data on GDP growth is released. We make the following observations: First, the root-mean-square forecast errors of all three models are on a slightly downward-sloping path throughout the forecasting period. This indicates that the data released over the forecasting period marginally improve the forecasting performance of the model. Second, the root-mean-square forecast errors of models 1 and 2 are nearly identical, which indicates that including the orthogonal financial factor into the model does not improve the ability to

Figure 9. Nowcast Evaluation



Source: Authors' calculations.

Note: Both charts show the values over the 2005:Q1–2019:Q3 sample, averaged over the distance to the release date of GDP.

forecast the mean of the growth distribution. Third, although the financial variables could in principle still help by providing timely information about the global factor, this contribution is only marginal, as is evident by comparing the root-mean-square forecast errors of models 1 and 2 (which use financial data) to model 3 (which does not). This is also apparent from the bottom chart, which shows the predictive scores of the three models. This measure accounts for the accuracy of the entire predictive distribution of GDP growth, not just the mean. Only an ever so slight improvement of the forecasting performance of models 1 and 2 (which use financial data) over model 3 (which does not) is noticeable.

SUMMARY Our out-of-sample test of the predictive ability of a nowcasting model in which we augment the standard global factor with an orthogonal financial factor reaches a disappointing conclusion: the performance of the model with both the global and financial factor is largely indistinguishable—in terms of root-mean-square forecast error and predictive score—from a model with only the global factor. The inclusion of financial variables into the global factor does lead to a small improvement in predictive score relative to a model with only a nonfinancial factor. This is probably due to the timeliness of financial variables, which can provide marginally earlier updates to the expected path of GDP growth at very short horizons.

III. How Does the Distribution of GDP Growth Change over Time?

The previous section demonstrated that there may be some limited out-of-sample information about the time-varying forecast distribution of GDP growth, although most of the predictive information comes from a global factor, not specifically financial variables. However, the method used there did not allow us to quantify the uncertainty surrounding any putative time variation in the conditional moments. In this section, we estimate a full statistical model of post-1975 US GDP growth that allows conditional moments to vary flexibly over time. Crucially, we will be able to quantify the uncertainty about the parameters in the model and thus the implied uncertainty about the evolution of the conditional moments of GDP growth. Unlike the previous section, we focus on in-sample results in this section. Thus, the only uncertainty is about the parameters of the model, which is assumed to be correctly specified. Even then, we find that the data are only informative about the conditional mean; the time variation of the conditional variance and higher moments is very imprecisely estimated. As a result, the time variation in the conditional recession probability and in the potential severity of recessions is driven almost exclusively by movements in the mean.

III.A. Data and Model

We model quarterly GDP growth as being driven by lagged GDP growth, as well as the global and financial factors estimated in section II. We use the final estimates of these factors. In this section we merely use these factors as a convenient set of low-dimensional explanatory variables, whereas the next section will attempt to attribute any explanatory power to individual variables with more direct economic interpretation. The sample period for estimation is 1975:Q2–2019:Q2. Online appendix S.E runs various benchmark linear forecast regressions using the global and financial factors. These benchmark regressions reveal that both factors potentially could contribute to the mean forecasts, at least in sample. However, we are primarily interested in going beyond the mean.

We assume that the one-quarter-ahead conditional distribution of GDP growth is given by the flexible skew *t*-distribution developed by Azzalini and Capitanio (2003). The distribution is indexed by four parameters: location μ, scale σ, shape α, and heavy-tailedness ν. These parameters influence—but do not directly equal—the conditional mean, variance, skewness, and kurtosis of the distribution. If $\alpha = 0$, the distribution reduces to the usual symmetric Student's *t*-distribution with ν degrees of freedom, which in turn reduces to the normal distribution when ν approaches ∞. If $\alpha > 0$, the distribution is positively skewed, while $\alpha < 0$ implies the opposite. Smaller values of ν correspond to fatter tails of the growth distribution (higher probability of abnormally low or high growth).

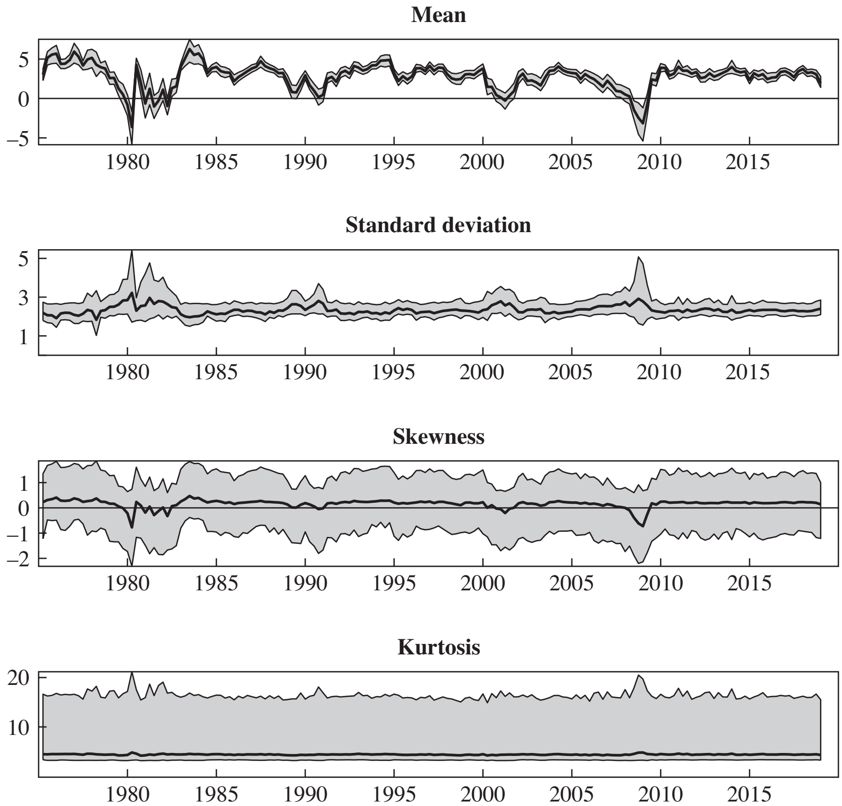
To allow the explanatory variables to influence several features of the GDP distribution, we model the location parameter $\mu = \mu_t$, the logarithm of the scale parameter $\log \sigma = \log \sigma_t$, and the shape parameter $\alpha = \alpha_t$ as being time-varying. These parameters are each assumed to depend linearly on an intercept, lagged GDP growth, and the lagged global and financial factors. The heavy-tailedness parameter ν is constant over time. This parameter mainly influences the kurtosis of the conditional growth distribution, and we will show below that there is little information in the data about time variation in higher moments anyway. We apply a Bayesian estimation procedure with weakly informative priors on the parameters.

The model and estimation procedure are described in detail in online appendix S.A. As discussed in the appendix, our model can be viewed as a fully Bayesian implementation of the estimation approach developed by Adrian, Boyarchenko, and Giannone (2019a) and used in section II. An advantage of our approach is that we can easily summarize the posterior uncertainty about time-varying parameters and moments.

III.B. Time Variation in US Moments and Tail Risk

Figure 10 shows that the data are only able to accurately pin down the time variation in the mean of the one-quarter-ahead conditional distribution of GDP growth. The standard deviation, skewness, and kurtosis of the forecast distribution are much less precisely estimated. The figure shows the posterior median and 90 percent credible interval for the moments at each point in time. The uncertainty is due to the fact that the underlying model parameters are estimated with varying degrees of precision in the post-1975 data. As is clear from the figure, the implied uncertainty about higher moments is large. Although the posterior median of the conditional

Figure 10. US Factor Model: Time-Varying Moments, One Quarter Ahead



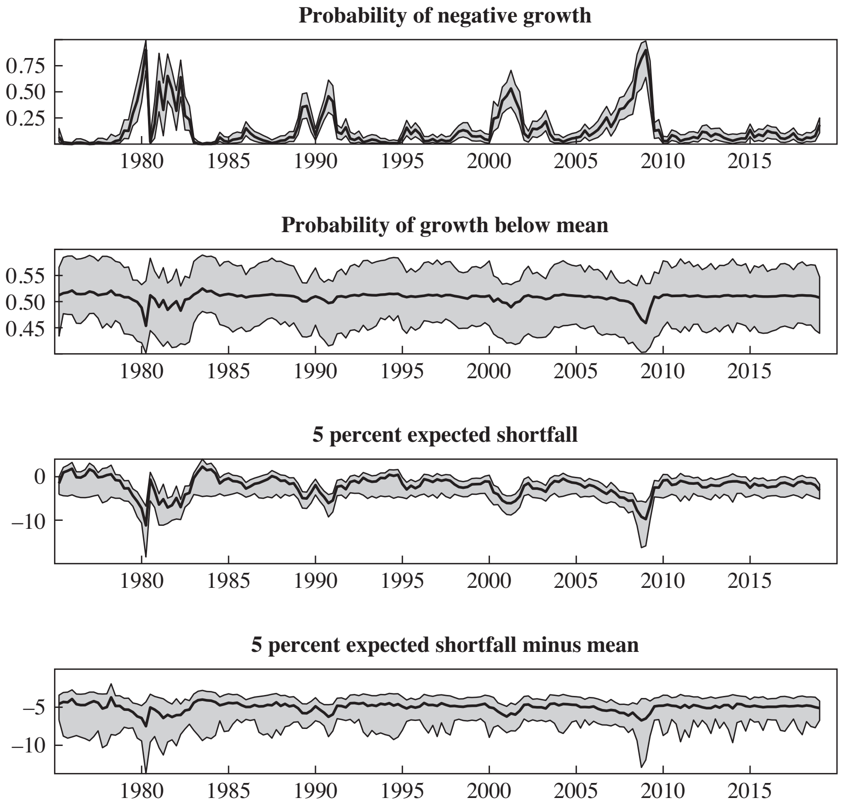
Sources: FRED-QD, FRED-MD, and authors' calculations.

Note: Time-varying moments of the one-quarter-ahead forecast distribution of GDP growth (annualized). The thick line is the posterior median (across parameter draws) at each point in time. The gray shaded band is the pointwise 90 percent posterior credible band (across parameter draws) at each point in time. The time axis shows the quarter in which the forecast is made.

standard deviation does fluctuate, quarters with potentially large swings are also associated with high uncertainty. The time paths of skewness and kurtosis are even more imprecisely estimated. Figure S.6 in online appendix S.F shows that all these results are qualitatively unchanged when we look at the conditional moments of the four-quarter-ahead forecast distribution.

How does the uncertainty about higher moments affect inferences about the left tail of the growth distribution? The top chart in figure 11 shows the time-varying implied one-quarter-ahead conditional probability of a

Figure 11. US Factor Model: Recession Probability and Expected Shortfall, One Quarter Ahead



Sources: FRED-QD, FRED-MD, and authors' calculations.

Note: Recession probability, probability of growth below the conditional mean, expected shortfall, and expected shortfall minus conditional mean for the one-quarter-ahead conditional distribution of GDP growth (annualized). The thick line is the posterior median (across parameter draws) at each point in time. The gray shaded band is the pointwise 90 percent posterior credible band (across parameter draws) at each point in time. The time axis shows the quarter in which the forecast is made.

recession (that is, negative growth in the following quarter). We see that the recession probability varies substantially over time and is reasonably precisely estimated. However, this is purely due to movements in the conditional mean of next-quarter GDP growth, as opposed to movements in the other moments. The second chart in the figure shows the conditional probability of GDP growth falling below the conditional mean; this probability does not vary much over time and is imprecisely estimated. The third chart in the figure shows the 5 percent expected shortfall, which is a measure of the severity of a recession should it materialize (specifically, it equals expected growth conditional on growth falling below the fifth percentile of its conditional distribution). The expected shortfall moves around over time, but the fourth chart—where the conditional mean has been subtracted—shows that this movement is almost entirely due to movement in the mean. We report analogous results for four-quarter-ahead forecasts in online appendix S.F; these are qualitatively similar.

Thus, there appears to be little exploitable time variation in the conditional GDP growth distribution apart from the mean. Although knowing the conditional standard deviation and higher moments would be very helpful for characterizing the risks to GDP growth, it appears that the available data for the United States are simply not sufficiently informative about these moments. On the positive side, movements in the conditional mean do appear to be partially predictable, at least in sample. Note that if we are interested in estimating the probability of recessions, and we shut down movement in all moments except for the mean, our model reduces to a probit forecasting model, which is a commonly used specification in applied work.

The financial factor contributes very little to the growth forecasts, whereas the global factor plays a larger role for the conditional mean. Online appendix S.F shows the posterior distribution of the model coefficients. The mean coefficients on both factors are statistically significant at conventional levels, but the coefficient on the global factor is estimated to be larger in magnitude. In the appendix we also investigate how the time-varying forecast moments shown in figure 10 change if we remove the global factor or the financial factor from the conditioning set when producing forecasts. Removing the financial factor has almost no discernible effect on any of the moments, whereas removing the global factor does lead to substantial changes in the path of the conditional mean, especially around the Great Recession period. Thus, as in the out-of-sample results in the previous section, the orthogonal financial factor plays a very minor role in short-term forecasting even in sample.

Figure 10 suggests that the unconditional skewness of US GDP growth is indistinguishable from zero, but this result masks a subtle feature of the posterior distribution of the underlying model parameters. In online appendix S.F we show that the marginal posterior distributions for the intercepts in the equations for the scale parameter σ_t and shape parameter α_t both exhibit a marked bimodality. These two parameters are highly negatively correlated in the posterior. In essence, the data cannot distinguish whether US GDP growth features either a low mean but positive skewness or a high mean but negative skewness. Notice that this is not a statement about variation in skewness *over time*, but simply a statement about posterior uncertainty about the nature of the unconditional GDP growth distribution. However, we show in online appendix S.F that if the model is estimated on the post-1980 sample, the positive skewness mode disappears. Figure 2 shows that US GDP growth was especially erratic in the late 1970s, and indeed growth from 1975 to 1979 has a positive sample skewness. Yet the post-1980 data point quite clearly toward negative unconditional skewness. We return to the estimation of unconditional skewness and kurtosis in section IV.

CROSS-COUNTRY EVIDENCE The fact that time variation in moments other than the mean is imprecisely estimated holds up in data for other OECD countries. We relegate the discussion of the cross-country data set to the next section, where these data are used more intensively. We estimate a global and financial factor separately for each of twelve other OECD countries, using the same method as we used for the United States. Online appendix S.F shows the estimated time-varying forecast moments for Australia, Italy, and Japan, which are representative of the other countries. In all cases, the conditional mean of GDP growth is estimated quite precisely, but posterior uncertainty about the model parameters translates into substantial uncertainty about the time paths of the conditional standard deviation, skewness, and kurtosis.

SUMMARY When using lagged GDP growth, a global factor, and a financial factor as predictors, it appears to be highly challenging to accurately estimate the time variation in the conditional variance, skewness, and kurtosis of GDP growth. The conditional mean, however, is reasonably precisely estimated, and it does appear to vary substantially over time. This is true in data for the United States and for other OECD countries. Hence, at least if we ignore out-of-sample forecasting issues, GDP growth forecasting is not a completely futile exercise at short horizons—though all the action is in the mean and none in the tails. More generally, our results demonstrate the importance of taking parameter uncertainty into account when making inferences about rare events from relatively short time series.

However, because we focused on factors as predictors, it remains a possibility that individual economic variables might provide strong signals about risks to GDP growth. We turn to this question in the next section.

IV. Which Variables Predict Growth Risk?

Do real activity and financial conditions indexes represent the best way to predict and describe growth vulnerability? Policymakers and academics alike may additionally be interested in which specific economic variables carry the most predictive power, for several reasons. First, when designing macroprudential policies or when explaining such policies to the public, it would be useful to know the most important economic predictor variables, narrowly defined. Second, financial indexes—such as the NFCI used by Adrian, Boyarchenko, and Giannone (2019a)—are usually not constructed to explicitly optimize the ability to forecast tail risk in GDP growth. Thus, it is possible that additional predictive power can be gleaned from considering predictor variables individually. Finally, detailed results on the performance of individual predictor variables may shine light on mechanisms that can guide theoretical model building.

In this section we complement the factor-based analysis of section III by performing a variable selection exercise to find those specific economic time series that best forecast various moments of GDP growth. We do this by estimating a conditional heteroskedasticity model and the dynamic skew-*t* model considered in the previous section on US and cross-country data sets, with a wide array of candidates for predictor variables. Rather than focusing directly on tail risks, we break down our results by the conditional moments of GDP growth, since this sheds more light on potential mechanisms. Our fully Bayesian approach allows us to describe the uncertainty surrounding the variable selection. For simplicity and clarity, we restrict attention to one-quarter-ahead forecasting in this section.

Relative to the literature, our contribution here is to select individual variables—among a large set of candidate variables—that predict GDP growth, its volatility, and higher moments, in data for the United States and for twelve other OECD countries. In contrast to the multicountry analyses of Adrian and others (2018) and Brownlees and Souza (2019), our focus is on variable selection and on characterizing cross-country heterogeneity in growth dynamics. Unlike these papers, we do not explore the role of the forecast horizon.

IV.A. Data

We employ two different data sets: a quarterly US data set and a multi-country data set for thirteen OECD countries. In addition to GDP growth (the outcome variable), both data sets contain an extensive set of possible predictor variables. The US data set is especially rich and extends back to 1975, while the predictors in the multicountry data set are slightly more limited in scope and extend back to 1980.

The quarterly US data set is based on the Federal Reserve Economic Data Quarterly Database (FRED-QD) data set constructed by Michael W. McCracken and Serena Ng, building on earlier work by Stock and Watson (2012).[8] This data set is frequently used for high-dimensional prediction in macroeconomics due to its broad scope, reliable data quality, and ease of availability. We select series from various categories of real, price, and financial variables. Though the selected financial series do not cover the full universe used to construct the NFCI, we do include corporate spreads; government bond yields; credit and loan volume; federal, corporate, and household balance sheet variables; stock price and dividends; implied volatility; and exchange rates. We supplement with data from Global Financial Data, Inc., and Haver Analytics on commodity prices; consumer, business, and purchasing manager surveys; and stock trading volume. This yields a total of forty-three predictor variables.

The multicountry data set covers thirteen OECD countries, with up to thirty-four predictor variables for each country. As in the US data described above, the potential predictor variables include a variety of real, price, survey, and financial variables. Our overarching goal is to ensure that variable definitions and samples are comparable across countries, so that any cross-country heterogeneity can be interpreted in a straightforward way. The thirteen countries are Australia (AUS), Belgium (BEL), Canada (CAN), Switzerland (CHE), Germany (DEU), Spain (ESP), France (FRA), the United Kingdom (GBR), Italy (ITA), Japan (JPN), the Netherlands (NLD), Sweden (SWE), and the United States (USA).[9] Our primary data source is the OECD Economic Outlook and Main Economic Indicators databases. We supplement with data from the BIS on house prices and credit, financial data from Global Financial Data, Inc., and household and business surveys from Haver Analytics.

Exploiting data from several countries could in principle ameliorate the inevitable data limitations when estimating the effect of financial indicators on real growth vulnerability (Adrian and others 2018). According to Carmen Reinhart's classification, the United States has undergone only

8. See the Federal Reserve Bank of St. Louis, Economic Research, https://research.stlouisfed.org/econ/mccracken/fred-databases/.

9. Adrian and others (2018) consider the same countries, excluding Belgium and the Netherlands.

two banking crises since 1980: the savings and loan crisis in the late 1980s and the global financial crisis of 2007–10.[10] However, every year from 1980 to 2014, with the exception of 2002–06, has witnessed a new or ongoing banking crisis in at least one of the thirteen countries in our data set. If we include currency crises in the calculation, only the years 2004 and 2006 were crisis-free in all thirteen countries. In an average year, 3.7 countries experience a crisis (standard deviation 2.7). From 1980 to 2016 there have been a total of ninety-nine country-years of banking crises and forty-seven country-years of currency crises for the countries in our data set (just nine country-years experienced both types of crisis at once).

The full list of all US and multicountry predictor variables (and their abbreviations) can be found in online appendix S.B.

To make coefficients comparable across different predictor variables, we standardize all predictors (but not GDP growth) to have sample mean zero and variance 1, separately for each country.

IV.B. Which Variables Forecast Growth and Its Volatility?

We first attempt to identify important predictors of the mean and volatility of GDP growth. We will initially restrict attention to a more parsimonious version of the dynamic skew-t model from section III. Specifically, we assume that only the mean and variance can vary over time, shutting down any potential time variation in higher moments. This conditional heteroskedasticity model was also analyzed by Adrian, Boyarchenko, and Giannone (2019a).

Because we are interested in selecting the relevant predictor variables among a large set of candidates, we employ a Bayesian prior distribution on the model parameters that imposes approximate sparsity, that is, it prefers parsimonious (and thus interpretable) models. Specifically, we impose the "horseshoe prior" of Carvalho, Polson, and Scott (2010), which essentially assumes that the coefficients on the various predictors are either relatively small or relatively large. The practical consequence of imposing this prior is that the posterior distribution will shrink many of the coefficients heavily toward zero, thus yielding a parsimonious model. However, the coefficients on those predictors that are most informative in the data will be shrunk very little. Since we continue to adopt a fully

10. See the data set collected by Carmen Reinhart and colleagues, Harvard Business School, Behavioral Finance and Financial Stability, https://www.hbs.edu/behavioral-finance-and-financial-stability/data/Pages/global.aspx.

Bayesian approach to inference, it is easy to quantify the uncertainty about the parameters in the model. We give further details about the estimation procedure in online appendix S.A.

RESULTS: US DATA We first estimate the model on the quarterly US data set from 1975:Q2 to 2019:Q2. Lagged GDP growth turns out not to be especially important for either the conditional mean or volatility, conditional on the other predictor variables discussed below. Hence, we report the results for the lagged growth coefficients and the intercepts in online appendix S.G.

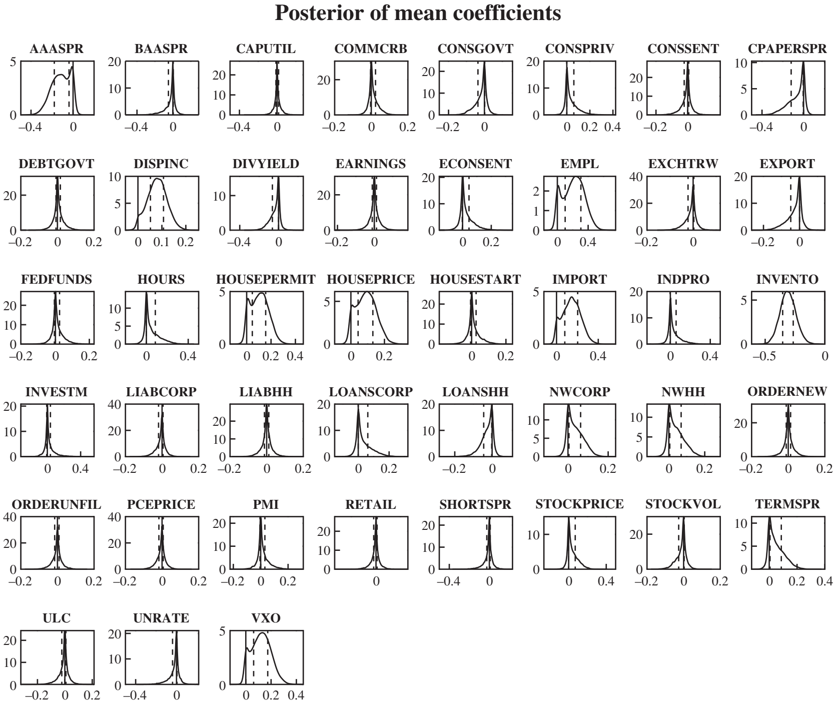
Mean forecasting. Which variables help predict the mean of GDP growth? Figure 12 shows the posterior densities for the mean predictor coefficients. Recall that all predictors have been standardized, so that the magnitudes of different coefficients are immediately comparable. About a third of the variables are found to have high posterior probability of being at least somewhat economically important. There is especially high posterior probability of inventories (INVENTO) being an economically important predictor of the mean of GDP growth, with statistically significant roles also played by disposable income (DISPINC), employment (EMPL), new housing permits (HOUSEPERMIT), house prices (HOUSEPRICE), and imports (IMPORT).

The only two financial variables that have a high probability of being important for the mean are implied volatility (VXO) and the spread between AAA corporate bonds and ten-year Treasuries (AAASPR). Perhaps surprisingly, the coefficient on the term spread (TERMSPR) is estimated to be small. There is only weak evidence that credit aggregates may play some role, although business loans (LOANSCORP), business net worth (NWCORP), and household net worth (NWHH) cannot be entirely ruled out.

Volatility forecasting. When it comes to volatility forecasting, there is strong evidence of predictive power for only a few variables. Figure 12 shows the posterior densities of the volatility coefficients. The coefficient on the AAA corporate bond spread (AAASPR) has substantial posterior mass at values in the range [–0.3, –0.1] (the posterior median is –0.16), indicating that a ceteris paribus one standard deviation increase in this spread is associated with a 10–30 percent increase in GDP growth volatility, a potentially substantial effect. Yet the bimodal nature of the posterior density reflects the fact that the data, combined with our prior belief in sparsity, cannot entirely rule out that even this coefficient may be close to zero.

None of the other predictor variables are unambiguously important for volatility forecasting. Other than the AAA spread and lagged GDP growth, no coefficient has a posterior median greater than 0.05 in magnitude. There are five other variables for which the posterior probability of

Figure 12. US Conditional Heteroskedasticity Model Posterior Densities

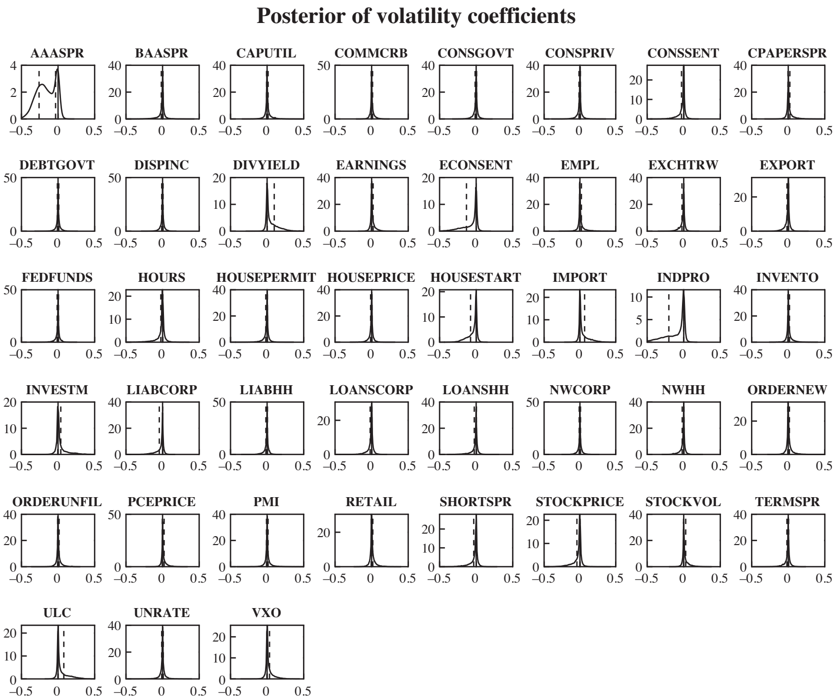


their coefficients exceeding 0.05, or being below –0.05, lies in the range 30–50 percent: business condition surveys (ECONSENT), housing starts (HOUSESTART), and industrial production (INDPRO) all possibly have a negative association with volatility, while the S&P 500 dividend yield (DIVYIELD) and unit labor cost index (ULC) possibly have a positive association with volatility. Of these variables, the one with the highest degree of posterior certainty is industrial production, for which the posterior probability of lying below –0.05 is a modest 48 percent.

RESULTS: CROSS-COUNTRY DATA Are the predictors of GDP growth and its volatility robustly identifiable across several developed countries? Estimating the conditional heteroskedasticity model separately on thirteen OECD countries from 1980:Q1 to 2018:Q4, we find that the answer to this question is a resounding no.

Mean forecasting. Although we found encouraging in-sample results on mean forecasting in US data, the precise identities of the relevant predictor variables appear to be highly heterogeneous across the thirteen

Figure 12. US Conditional Heteroskedasticity Model Posterior Densities *(Continued)*



Sources: FRED-QD; Global Financial Data, Inc.; Haver Analytics; and authors' calculations.

Note: Posterior densities of the coefficients on mean and variance predictor variables in the conditional heteroskedasticity model. Vertical dashed lines indicate posterior interquartile ranges. A coefficient value of 0.1 means that an increase in the predictor by one standard deviation is associated with a 0.1 percentage point increase in the conditional mean of quarter-on-quarter GDP growth (left, p. 32) or with a 10 percent increase in the conditional volatility of quarter-on-quarter GDP growth (right, above).

OECD countries. Table 2 shows summary statistics of the posterior distributions of the mean predictor coefficients across countries. Other than lagged GDP growth, only the national stock index (STOCKPRICE) is significant at the 50 percent level for more than half the countries (in the sense that the posterior interquartile range excludes zero). The coefficients on consumer sentiment (CONSSENT) and the manufacturing production index (MANUF) also have posterior probability greater than 20 percent (on average across countries) of being larger than 0.1, meaning that a one standard deviation increase is associated with 10 basis points higher quarter-on-quarter GDP growth. Other than the stock index, no other financial variables seem important for more than a few countries, including various financial spreads and credit aggregates.

Table 2. Cross-Country Conditional Heteroskedasticity Model: Posterior of Mean Coefficients

		Average across countries			
Variable	*#*[a]	*Median*[b]	*Signif*[c]	*P > .1*[d]	*P < –.1*[d]
CA	13	–0.0006	0.08	0.01	0.01
COMMCRB	13	0.0055	0.15	0.03	0.00
CONSGOVT	13	–0.0054	0.08	0.00	0.03
CONSPRIV	13	0.0289	0.23	0.15	0.00
CONSSENT	7	0.0245	0.43	0.17	0.00
CREDCORP	13	0.0019	0.08	0.03	0.02
CREDCORPBNK	13	–0.0052	0.08	0.02	0.04
CREDHH	12	0.0024	0.00	0.04	0.01
DIVYIELD	13	–0.0178	0.31	0.01	0.11
ECONSENT	6	0.0067	0.33	0.06	0.01
EMPL	13	0.0296	0.31	0.15	0.00
EXCHEFF	13	–0.0003	0.00	0.01	0.01
EXCHUSD	12	–0.0081	0.08	0.02	0.05
EXPORT	13	0.0063	0.08	0.05	0.01
GDPDEF	13	0.0010	0.15	0.01	0.01
HOURS	12	0.0126	0.08	0.07	0.00
HOUSEPERMIT	6	0.0261	0.33	0.14	0.00
HOUSEPRICE	13	0.0211	0.46	0.11	0.00
HOUSESTART	8	0.0102	0.13	0.06	0.01
IMPORT	13	0.0155	0.23	0.10	0.00
INTRBNKRATE	13	0.0003	0.00	0.01	0.01
INVESTM	13	0.0227	0.38	0.15	0.03
MANUF	13	0.0497	0.38	0.21	0.00
PMI	1	0.0079	0.00	0.07	0.00
RETAIL	12	0.0011	0.17	0.02	0.02
STOCKPRICE	13	0.0352	0.54	0.20	0.00
STOCKRV	13	–0.0007	0.00	0.01	0.02
STOCKVOL	10	0.0081	0.20	0.06	0.00
TERMSPR	13	0.0072	0.23	0.05	0.01
TERMTRADE	13	0.0032	0.08	0.02	0.01
ULC	12	0.0010	0.25	0.05	0.02
UNRATE	13	–0.0103	0.23	0.00	0.08
VXO	13	0.0015	0.00	0.01	0.01
YIELDSPRUS	12	–0.0039	0.08	0.00	0.03
YLAG	13	0.1449	0.77	0.59	0.13

Sources: OECD; BIS; Global Financial Data, Inc.; Haver Analytics; and authors' calculations.
Note: Summary statistics of the mean coefficient posterior distributions for thirteen OECD countries.
a. Number of countries present in the data.
b. Posterior median of coefficient.
c. Indicator for whether posterior interquartile range for coefficient excludes zero.
d. Posterior probability that coefficient is > .1 or < –.1, respectively.

Volatility forecasting. Cross-country heterogeneity is even more pervasive in volatility forecasting. Table 3 shows summary statistics of the posterior distributions of the volatility predictor coefficients across countries. The only volatility predictor variable that is significant at the 50 percent level for more than five countries is the term spread (TERMSPR). Turning to economic significance, it is only the coefficients on S&P 100 implied volatility (VXO, an international variable) and on lagged GDP growth itself (YLAG) that have a nonnegligible posterior probability of being larger than 0.05 in magnitude for more than a handful of countries. Recall that a coefficient magnitude of 0.05 means that a one standard deviation change in the variable predicts a 5 percent change in volatility, a modest amount.

Very few of the posterior medians of the volatility coefficients are economically significant, as shown in figure 13. The only three variables whose posterior medians are large in magnitude for two or more countries are stock prices (STOCKPRICE), S&P 100 implied volatility (VXO), and the ten-year government bond spread vis-à-vis the US (YIELDSPRUS). However, with the exception of VXO, the signs of the estimated effects of these variables differ across countries. If interest centers on specific countries, however, we do find strong evidence of substantial predictive power for a small number of additional variables, such as economic sentiment surveys (ECONSENT) and the term spread (TERMSPR) for the Netherlands, and house prices (HOUSEPRICE) for Japan.

SUMMARY We arrive at a negative conclusion: though it is possible to find strong evidence of a few important mean predictors and (less frequently) volatility predictors for individual countries—such as for the United States—generalizing to other countries seems fraught with danger. There is little agreement across countries about the identity and sign of important mean and volatility predictors, despite our efforts to construct a data set with comparable variable definitions and data availability.

Contrary to the conjecture mentioned in section I that financial spreads and credit aggregates might carry different information about growth vulnerability, we do not find a robust role for either type of variable in mean or volatility forecasting. No financial variable in our data set plays a statistically and economically significant role in forecasting GDP growth at short horizons for more than a handful of the thirteen countries we consider. We stress, though, that our cross-country data set does not contain a measure of corporate borrowing spreads due to data availability. Thus, our analysis does not overturn the existing literature discussed in the introduction, although it does caution against putting too much faith in single-country analyses.

Table 3. Cross-Country Conditional Heteroskedasticity Model: Posterior of Volatility Coefficients

		Average across countries			
Variable	*#*[a]	*Median*[b]	*Signif*[c]	*P > .05*[d]	*P < –.05*[d]
CA	13	–0.0073	0.31	0.06	0.17
COMMCRB	13	–0.0134	0.23	0.07	0.18
CONSGOVT	13	0.0031	0.00	0.11	0.05
CONSPRIV	13	0.0012	0.15	0.13	0.11
CONSSENT	7	–0.0159	0.14	0.04	0.27
CREDCORP	13	–0.0013	0.00	0.07	0.12
CREDCORPBNK	13	–0.0059	0.08	0.06	0.14
CREDHH	12	–0.0012	0.00	0.07	0.12
DIVYIELD	13	0.0018	0.00	0.11	0.05
ECONSENT	6	–0.0745	0.33	0.04	0.32
EMPL	13	–0.0010	0.00	0.08	0.11
EXCHEFF	13	0.0039	0.08	0.11	0.08
EXCHUSD	12	–0.0007	0.00	0.08	0.10
EXPORT	13	–0.0019	0.00	0.05	0.10
GDPDEF	13	–0.0008	0.08	0.07	0.07
HOURS	12	0.0073	0.25	0.13	0.11
HOUSEPERMIT	6	–0.0076	0.17	0.07	0.17
HOUSEPRICE	13	0.0064	0.23	0.11	0.14
HOUSESTART	8	–0.0051	0.25	0.06	0.14
IMPORT	13	0.0079	0.08	0.15	0.05
INTRBNKRATE	13	–0.0014	0.08	0.09	0.09
INVESTM	13	0.0017	0.00	0.09	0.08
MANUF	13	–0.0107	0.15	0.07	0.16
PMI	1	–0.0008	0.00	0.04	0.09
RETAIL	12	–0.0021	0.17	0.10	0.10
STOCKPRICE	13	0.0019	0.23	0.11	0.16
STOCKRV	13	0.0025	0.08	0.13	0.07
STOCKVOL	10	0.0010	0.20	0.11	0.11
TERMSPR	13	–0.0379	0.54	0.04	0.31
TERMTRADE	13	0.0106	0.15	0.14	0.07
ULC	12	0.0057	0.17	0.15	0.05
UNRATE	13	0.0106	0.08	0.15	0.07
VXO	13	0.0596	0.38	0.40	0.01
YIELDSPRUS	12	0.0321	0.42	0.25	0.12
YLAG	13	–0.0283	0.38	0.34	0.42

Sources: OECD; BIS; Global Financial Data, Inc.; Haver Analytics; and authors' calculations.

Note: Summary statistics of the volatility coefficient posterior distributions for the thirteen OECD countries.

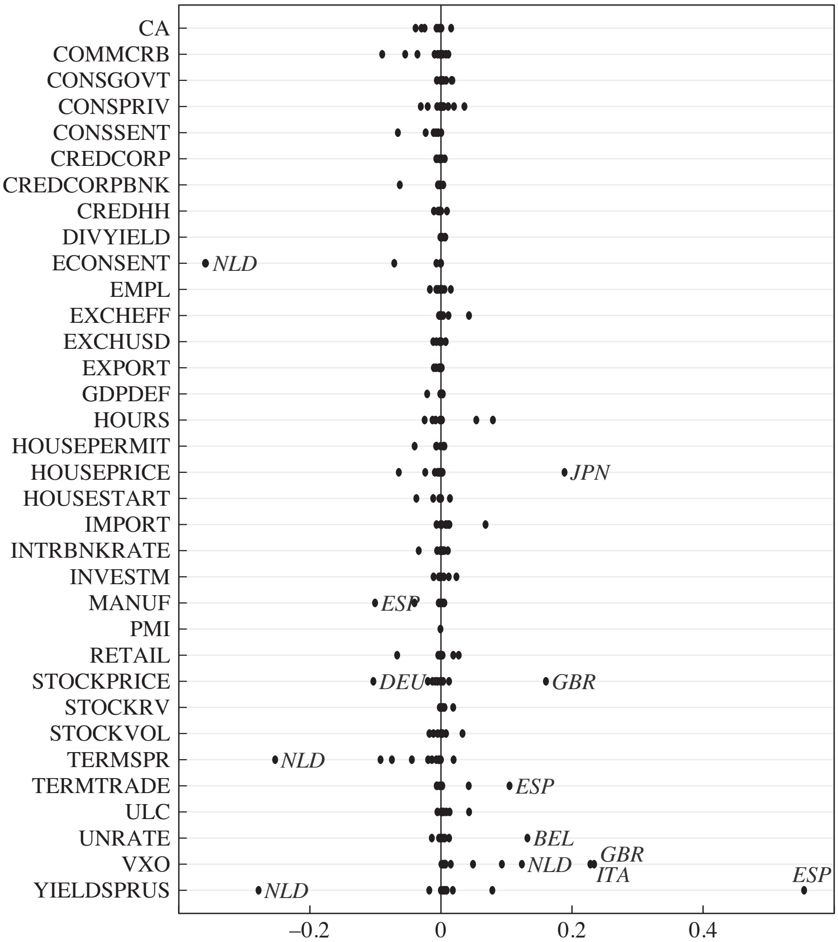
a. Number of countries present in the data.

b. Posterior median of coefficient.

c. Indicator for whether posterior interquartile range for coefficient excludes zero.

d. Posterior probability that coefficient is > .05 or < –.05, respectively.

Figure 13. Cross-Country Conditional Heteroskedasticity Model: Posterior Medians of Volatility Coefficients



Sources: OECD; BIS; Global Financial Data, Inc.; Haver Analytics; and authors' calculations.
Note: Each row in the plot corresponds to a variable, while the dots in each row correspond to different countries.

IV.C. Which Variables Are Informative about Higher Moments?

Can we go beyond the mean or volatility and characterize the predictors of time variation in the skewness of GDP growth? To answer this question, we turn again to the full dynamic skew-*t* model described in section III, but instead of using a small number of factors as explanatory variables, we use our full set of individual economic predictor variables.[11]

In short, we find little robust evidence of individual predictors being informative about the time variation of skewness. In online appendix S.G we define a measure of the skewness of the forecast distribution with interpretable units, called "TVD." This measure lies between zero and 1, with 1 indicating substantial skewness and zero indicating a symmetric distribution. Using this measure, we find that no predictor variable has an economically significant positive or negative effect on the time variation of skewness in more than a few of the countries in our analysis. The results are relegated to the appendix due to space constraints.

The distribution of GDP growth does exhibit clear unconditional skewness as well as moderate kurtosis in many countries. Table 4 displays, for each country, posterior summaries of α_t, TVD(α_t), and ν (Japan and Spain have been dropped from the analysis due to numerical convergence issues for these countries). Based on time-averaged TVD, most countries exhibit substantial skewness, as values of TVD around 25–40 percent indicate substantial departures from symmetry. From the time-averaged α_t values it is clear, however, that the direction of skewness varies across countries: GDP growth tends to be negatively skewed in Switzerland, Germany, France, the Netherlands, and the United States, and positively skewed in the other countries. As expected based on the above results, there does not appear to be substantial time variation in the extent of the skew, as can be seen by comparing the average and standard deviation of TVD over time within countries.[12] As for kurtosis, all countries but the United Kingdom have posterior medians of ν in excess of 10, indicating at most moderately fat tails.

SUMMARY Skewness—and to a lesser extent fat tails—do seem to be pervasive features of the unconditional GDP growth distribution in many

11. It turns out to be computationally difficult to impose a prior belief in sparsity in the full dynamic skew-*t* model, unlike in the conditional heteroskedasticity model considered in section IV.B. Hence, we here instead use conventional normal shrinkage priors. See online appendix S.A for details.

12. This is consistent with the conclusion of Adrian, Boyarchenko, and Giannone (2019a, 1276), who however do not report measures of parameter uncertainty.

Table 4. Cross-Country Skew-*t* Model: Unconditional Skewness and Kurtosis

Country	*Avg*(α)[a]	*Avg(TVD)*[b]	*Std(TVD)*[b]	*Q1*(ν)[c]	*Med*(ν)[c]	*Q3*(ν)[c]
AUS	5.224	0.387	0.087	12.0	18.0	26.7
BEL	1.747	0.311	0.092	7.6	13.0	21.6
CAN	0.472	0.273	0.101	12.0	18.3	27.4
CHE	–0.821	0.243	0.081	8.5	13.0	20.2
DEU	–5.574	0.363	0.093	13.5	20.1	29.4
FRA	–0.160	0.248	0.100	12.0	18.2	26.9
GBR	1.578	0.307	0.107	4.5	7.1	12.5
ITA	4.229	0.369	0.089	12.8	19.4	28.5
NLD	–4.719	0.392	0.087	10.9	16.7	25.4
SWE	2.381	0.331	0.114	6.5	10.0	16.2
USA	–2.194	0.321	0.096	14.6	21.5	31.0

Sources: OECD; BIS; Global Financial Data, Inc.; Haver Analytics; and authors' calculations.
Note: Unconditional higher moments of the GDP growth distribution, for eleven OECD countries.
a. Posterior mean of average (across time) of α_t.
b. Posterior means of average and standard deviation (across time) of TVD(α_t), respectively.
c. Posterior first quartile, median, and third quartile of ν, respectively.

countries, but attributing the time variation in these higher moments to specific interpretable economic variables appears challenging given available data. This echoes the result in section III, which used aggregated factors as predictor variables. In particular, corporate or household credit growth is not robustly associated with negative conditional skewness of GDP growth. Adrian and others (2018) find evidence for an interaction effect in cross-country data: when credit growth is high, financial conditions are stronger predictors of risks to GDP growth at short horizons. Although we do not have explicit interaction terms in our model, the dynamic skew-*t* model can in principle generate this empirical pattern if credit growth negatively affects skewness while other financial variables affect the mean or variance of GDP growth. However, we do not find evidence for this mechanism in our data set. It is an interesting topic for future research to extend the dynamic skew-*t* model to allow for further state dependence.

V. Case Study: COVID-19

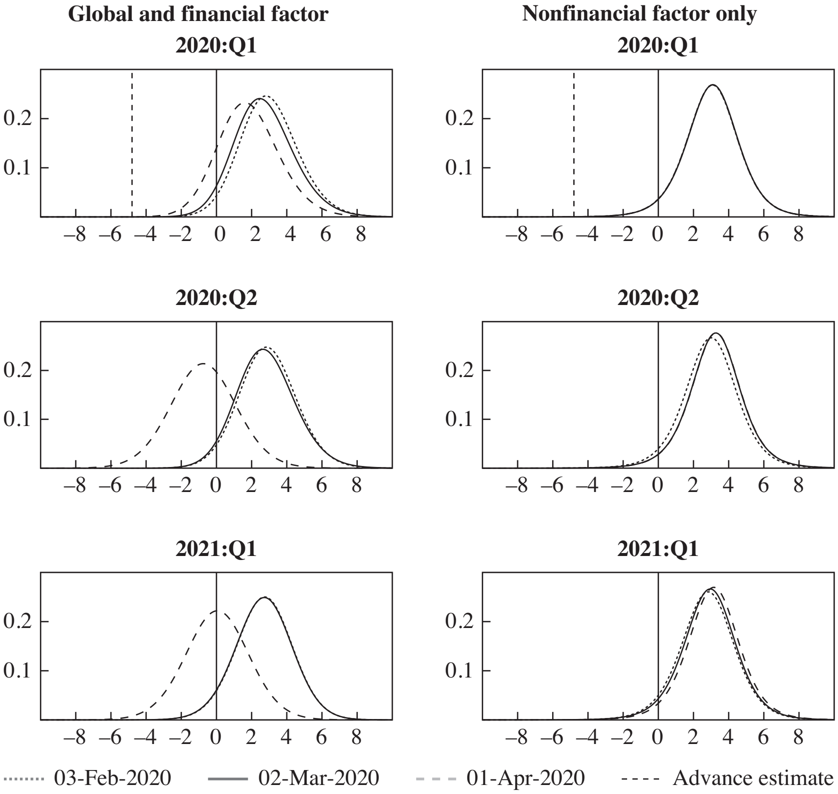
COVID-19 struck the world economy unexpectedly. A sharp recession in the United States, as in other parts of the world, was induced by the lockdown of a large part of the economy. Given the typical delay of macroeconomic information, it has been very difficult for traditional nowcasting and forecasting models to obtain meaningful numbers for the evolution of GDP in the first and second quarters of 2020. The most recent published figure for first quarter growth is −4.8 percent, well below expectations. This provides a natural experiment for the analysis of this paper. Would the nowcast in real time have been more accurate in models that include financial variables?

Using the same nonparametric real-time estimation approach as in section II, we compute here the predictive distribution of GDP for the first and second quarters of 2020 and the first quarter of 2021. We condition on information available at three different dates: the first business days of February, March, and April 2020. It is important to notice that—apart from financial variables—no common business cycle indicators relating to the lockdown period were available until the end of April. However, news stories and policy discussion of the pandemic were rampant starting in January 2020, and this information could potentially have been reflected in asset prices, business and consumer surveys, and so on. We consider two models. The first includes the macrofinancial common factor (global factor) and the orthogonal financial factor (results are shown in figure 14, left side). The second model conditions on the nonfinancial factor only (figure 14, right side).

Figure 14 shows that financial variables do provide useful timely information about the COVID-19 downturn, although they react relatively late and severely undershoot the magnitude. The forecast distributions of GDP growth for 2020:Q1, 2020:Q2, and 2021:Q1 hardly move at all if we condition only on lagged GDP growth and the nonfinancial factor, even though in reality the economy contracted markedly in March. However, when conditioning on the global and financial factors, the predictive distributions for the first two quarters of 2020 and for one year ahead start moving to the left in the beginning of April. According to our data release calendar, and given the ad hoc convention that financial variables for March are released on April 1, at that date the only information available concerning March was financial data. At that time, surveys and macro variables were available only for January and February, before the lockdown went into effect. Thus, financial variables proved to be useful for nowcasting this particular episode. Notice, however, that none of the forecasts came close to predicting the actual scale of the downturn. Moreover, financial variables only started flashing warning signs in late February, mere days before dramatic policy actions were introduced in several US states.

Why did financial variables not similarly provide a timely warning in the early stages of the 2008–09 recession, as discussed in section II? The difference is that in January 2009 when the model updated the estimate for

Figure 14. Predictive Distributions of GDP Growth in the COVID-19 Crisis



Sources: Bureau of Economic Analysis (BEA) and authors' calculations.
Note: Quarter-by-quarter evolution of the predictive distributions for the COVID-19 crisis, for the models including the global factor and the financial factor (left side), and the nonfinancial factor (right side). The charts for 2020:Q1 also report the BEA advance estimate of annualized GDP growth.

2008:Q4, it could exploit information from both macro and financial data for October and November. These data points already signaled a fall in output. Hence, in this case the information from financial variables about December 2008 just served to confirm the negative signal, without providing truly novel information, unlike in the COVID-19 episode. In summary, this small COVID-19 case study suggests that financial variables can sometimes be useful timely indicators at short horizons when no other information is available from macroeconomic surveys and the like. Moreover, while financial variables correctly hinted at a directional movement in the GDP growth distribution, the actual forecast was still very poor relative to the realization. Thus, the conclusion of our analysis of the uncertainty surrounding forecasts of moments other than the mean, which we have provided in the previous sections, remains in force: one should not place too much confidence in the signaling ability of financial variables.

VI. Summary and Conclusions

The results presented in this paper indicate that financial variables have very limited predictive power for the distribution of GDP growth at short horizons, especially—but not limited to—the tail risk. Two factors drive these results.

First, moments other than the mean are estimated very imprecisely. Although our findings confirm that GDP growth in many countries exhibits a skewed unconditional distribution, it is very hard to precisely estimate the dynamics over time of variance, skewness, and kurtosis conditional on financial and macroeconomic variables. This implies that, when computing the probability of recessions from the estimated moments, we essentially obtain what we could have obtained by using a probit model. These results are true whether we allow individual variables to enter the model in a flexible way via a variable selection algorithm or we aggregate them as factors. The variable selection exercise does not point to any stable stylized facts, except for the finding that real indicators are selected more often than financial ones. While our results do not rule out a transmission of shocks from the level of variables to their variance and other moments, as sometimes postulated in stochastic volatility models, this mechanism is empirically tenuous.

Second, information in monthly financial variables is highly correlated with information in macroeconomic variables, especially in recessions, but the correlation is contemporaneous. As the economy enters a recession and we observe a fall in output, markets have a sudden change in sentiments which leads to a spike in the spread variables. A common factor extracted from financial and macro data usually predicts a fall in the mean of GDP during the onset of the recession, but no further predictive power is gained by adding an extra orthogonal factor capturing financial-specific information.

In our real-time nowcasting exercise, which takes into account data uncertainty and the release calendar of economic data, we showed that the timeliness advantage of financial variables is generally minuscule. The case study of the COVID-19 lockdown episode, however, shows that financial variables can in some unique instances provide early warning signs when other macroeconomic data are not yet available. Still, even in this episode, models with financial data missed the severity of the downturn. Thus, the timeliness of financial information may help in real time but should not be overinterpreted, and financial markets do not seem to contain much forward-looking information about the macroeconomy beyond the current quarter.

The substantial cross-country heterogeneity in the identities of important predictor variables calls for humility in theoretical model building: the precise channels of the financial-real vulnerability nexus are difficult to tease out from the available data. In particular, it is likely a mistake to treat broad financial conditions indexes as catchall representations of any arbitrary financial friction that is of theoretical interest. Lack of predictive power might be the result of time instability between financial variables and GDP, which in turn may be caused by changes to the financial system and the conduct of monetary policy. This is something to be investigated further in future research.

Future research may also investigate whether our methods overlook state dependency and interactions between financial fragility and macroeconomic dynamics. For example, Krishnamurthy and Muir (2017) find that the interaction between credit spreads and precrisis credit growth can forecast the severity of the crisis. Aikman and others (2016) find that when private nonfinancial leverage is above trend, an easing of financial conditions predicts an economic expansion in the near term and a contraction in the following quarters. This is an interesting line of research which has implication for policy, as emphasized by Adrian and others (2018). It implies that although recessions are fundamentally unpredictable, prudential action can make the system less fragile so that, when they occur, the damage is limited. Although we do not directly investigate the role of such interactions, our results at the very least suggest that empirical analysis of this phenomenon must be fraught with substantial estimation uncertainty.

ACKNOWLEDGMENTS We are grateful for very insightful comments from Mark Gertler, Domenico Giannone, Nellie Liang, Jim Stock, and seminar participants at NYU. Plagborg-Møller acknowledges that this material is based upon work supported by the NSF under Grant #1851665. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone. 2019a. "Vulnerable Growth." *American Economic Review* 109, no. 4: 1263–89.

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone. 2019b. "Multimodality in Macro-Financial Dynamics." Staff Report 903. New York: Federal Reserve Bank of New York. https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr903.pdf.

Adrian, Tobias, Federico Grinberg, Nellie Liang, and Sheheryar Malik. 2018. "The Term Structure of Growth-at-Risk." Working Paper 18/180. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2018/08/02/The-Term-Structure-of-Growth-at-Risk-46150.

Aikman, David, Andreas Lehnert, Nellie Liang, and Michele Modugno. 2016. "Financial Vulnerabilities, Macroeconomic Dynamics, and Monetary Policy." Finance and Economics Discussion Series 2016-055. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/econresdata/feds/2016/files/2016055pap.pdf.

Azzalini, Adelchi, and Antonella Capitanio. 2003. "Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew *t* Distribution." *Journal of the Royal Statistical Society: Statistical Methodology Series B* 65, no. 2: 367–89.

Basel Committee for Banking Supervision. 2010. *Guidance for National Authorities Operating the Countercyclical Capital Buffer*. Basel: Bank for International Settlements. https://www.bis.org/publ/bcbs187.pdf.

Bernanke, Ben S. 2018. "The Real Effects of the Financial Crisis." *Brookings Papers on Economic Activity*, Fall, 251–342.

Bernanke, Ben S., and Mark Gertler. 1989. "Agency Costs, Net Worth, and Business Fluctuations." *American Economic Review* 79, no. 1: 14–31.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." In *Handbook of Macroeconomics*, edited by J. B. Taylor and M. Woodford. London: Elsevier.

Bloom, Nicholas. 2014. "Fluctuations in Uncertainty." *Journal of Economic Perspectives* 28, no. 2: 153–76.

Bloom, Nicholas, F. Guvenen, and S. Salgado. 2019. "Skewed Business Cycles." Working Paper 26565. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26565.

Borio, C., and P. Lowe. 2002. "Asset Prices, Financial and Monetary Stability: Exploring the Nexus." Working Paper 114. Basel: Bank for International Settlements. https://www.bis.org/publ/work114.htm.

Brave, Scott, and Andrew Butters. 2012. "Diagnosing the Financial System: Financial Conditions and Financial Stress." *International Journal of Central Banking* 8, no. 2: 191–239.

Brownlees, Christian T., and Andre B. M. Souza. 2019. "Backtesting Global Growth-at-Risk." Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3461214

Brunnermeier, Markus K., Darius Palia, Karthik A. Sastry, and Christopher A. Sims. 2019. "Feedbacks: Financial Markets and Economic Activity." Working Paper 257. Princeton, N.J.: Princeton University, Center for Economic Policy Studies. https://gceps.princeton.edu/wp-content/uploads/2019/09/257_Sims-Brunnermeier.pdf.

Carvalho, Carlos M., Nicholas G. Polson, and James G. Scott. 2010. "The Horseshoe Estimator for Sparse Signals." *Biometrika* 97, no. 2: 465–80.

Claessens, Stijn, M. Ayhan Kose, and Marco E. Terrones. 2012. "How Do Business and Financial Cycles Interact?" *Journal of International Economics* 87, no. 1: 178–90.

Doz, Catherine, Domenico Giannone, and Lucrezia Reichlin. 2012. "A Quasi-Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models." *Review of Economics and Statistics* 94, no. 4: 1014–24.

Drehmann, Mathias, Claudio Borio, and Kostas Tsatsaronis. 2012. "Characterising the Financial Cycle: Don't Lose Sight of the Medium Term!" Working Paper 380. Basel: Bank for International Settlements. https://www.bis.org/publ/work380.htm.

Fagiolo, Giorgio, Mauro Napoletano, and Andrea Roventini. 2008. "Are Output Growth-Rate Distributions Fat-Tailed? Some Evidence from OECD Countries." *Journal of Applied Econometrics* 23, no. 5: 639–69.

Figueres, Juan Manuel, and Marek Jarociński. 2020. "Vulnerable Growth in the Euro Area: Measuring the Financial Conditions." *Economics Letters* 191.

Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. 2003. "Do Financial Variables Help Forecasting Inflation and Real Activity in the Euro Area?" *Journal of Monetary Economics* 50, no. 6: 1243–55.

Gertler, Mark, and Simon Gilchrist. 2018. "What Happened: Financial Factors in the Great Recession." *Journal of Economic Perspectives* 32, no. 3: 3–30.

Giannone, Domenico, Michele Lenza, and Lucrezia Reichlin. 2019. "Money, Credit, Monetary Policy, and the Business Cycle in the Euro Area: What Has Changed since the Crisis?" *International Journal of Central Banking* 15, no. 5: 137–73.

Giannone, Domenico, Lucrezia Reichlin, and David Small. 2008. "Nowcasting: The Real-Time Informational Content of Macroeconomic Data." *Journal of Monetary Economics* 55, no. 4: 665–76.

Giglio, Stefano, Bryan Kelly, and Seth Pruitt. 2016. "Systemic Risk and the Macroeconomy: An Empirical Evaluation." *Journal of Financial Economics* 119, no. 3: 457–71.

Gilchrist, Simon, and Egon Zakrajšek. 2012. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review* 102, no. 4: 1692–1720.

Greenspan, Alan. 2004. "Risk and Uncertainty in Monetary Policy." *American Economic Review* 94, no. 2: 33–40.

Hatzius, Jan, Peter Hooper, Frederic Mishkin, Kermit L. Schoenholtz, and Mark W. Watson. 2010. *Financial Conditions Indexes: A Fresh Look after the Financial Crisis*. Report presented at the Chicago Booth Initiative on Global Markets 2010 US Monetary Policy Forum, New York, February 26.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2011. "Financial Crises, Credit Booms, and External Imbalances: 140 Years of Lessons." *IMF Economic Review* 59, no. 2: 340–78.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2013. "When Credit Bites Back." *Journal of Money, Credit and Banking* 45, no. S2: 3–28.

Kiley, Michael T. 2018. "Unemployment Risk." Finance and Economics Discussion Series 2018-067. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/econres/feds/files/2018067pap.pdf.

Kilian, Lutz, and Simone Manganelli. 2008. "The Central Banker as a Risk Manager: Estimating the Federal Reserve's Preferences under Greenspan." *Journal of Money, Credit and Banking* 40, no. 6: 1103–29.

Kiyotaki, Nobuhiro, and John Moore. 1997. "Credit Cycles." *Journal of Political Economy* 105, no. 2: 211–48.

Krishnamurthy, Arvind, and Tyler Muir. 2017. "How Credit Cycles across a Financial Crisis." Working Paper 23850. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w23850.

Loria, Francesca, Christian Matthes, and Donghai Zhang. 2019. "Assessing Macroeconomic Tail Risk." Working Paper 19-10. Federal Reserve Bank of Richmond. https://www.richmondfed.org/publications/research/working_papers/2019/wp_19-10.

McCracken, Michael W., and Serena Ng. 2016. "FRED-MD: A Monthly Database for Macroeconomic Research." *Journal of Business and Economic Statistics* 34, no. 4: 574– 89.

Prasad, Ananthakrishnan, Selim Elekdag, Phakawa Jeasakul, Romain Lafarguette, Adrian Alter, Alan Xiaochen Feng, and Changchun Wang. 2019. "Growth at Risk: Concept and Application in IMF Country Surveillance." Working Paper 19/36. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2019/02/21/Growth-at-Risk-Concept-and-Application-in-IMF-Country-Surveillance-46567.

Reichlin, Lucrezia, Giovanni Ricco, and Thomas Hasenzagl. 2020. "Financial Variables as Predictors of Real Growth Vulnerability." Working Paper 05/2020. Deutsche Bundesbank. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3556506.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. "The Aftermath of Financial Crises." *American Economic Review* 99, no. 2: 466–72.

Schularick, Moritz, and Alan M. Taylor. 2012. "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008." *American Economic Review* 102, no. 2: 1029–61.

Stock, James H., and Mark W. Watson. 2003. “Forecasting Output and Inflation: The Role of Asset Prices.” *Journal of Economic Literature* 41, no. 3: 788–829.

Stock, James H., and Mark W. Watson. 2012. “Disentangling the Channels of the 2007–09 Recession.” *Brookings Papers on Economic Activity*, Spring, 81–135.

Valencia, Fabian, and Luc Laeven. 2012. “Systemic Banking Crises Database: An Update.” Working Paper 12/163. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Systemic-Banking-Crises-Database-An-Update-26015.

Comments and Discussion

COMMENT BY

MARK GERTLER This impressive paper examines the forecasting power of financial indicators for distribution of GDP growth. The motivation for examining the distribution of GDP growth is that most major financial crises feature sharp nonlinear contractions in GDP. The question then arises as to whether financial indicators can provide an early warning of these economic disasters. Indeed, important recent work by Adrian, Boyarchenko, and Giannone (2019) provides some hope that this may be the case. This work presents evidence that financial variables have predictive power for the lower quantile of GDP growth, which the literature has termed "growth at risk."

Perhaps not surprisingly the present paper provides a "forecaster's perspective." There is considerable emphasis on the statistical significance of the forecast. In addition, the authors focus on the marginal information that comes from financial variables, as I discuss shortly. Given the tough standards they apply, the authors show convincingly that financial variables do not provide significant marginal predictive power for the distribution of GDP growth.

In particular, the authors begin with time series of a large number of real and financial variables. They then proceed to construct two factors. The first is a global factor (GF) that characterizes common movements among the entire set of real and financial variables. Then from the financial variables alone they construct a financial factor (FF). The FF captures common movements in the financial variables that are orthogonal to the GF. In this respect, the FF isolates the marginal information from financial factors.

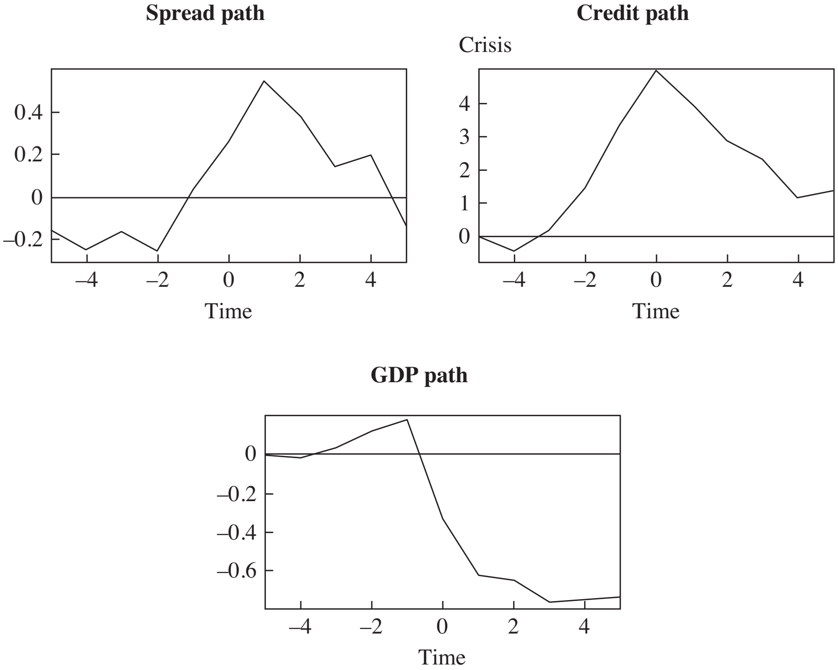
The authors then proceed to assess the predictive power of the FF for the distribution of GDP growth. Here they analyze both out-of-sample forecasting and in-sample parameter uncertainty. There are two main findings. First, it is difficult to predict moments other than the mean, even with the GF. Second, the FF adds little to the forecast. As a check against the possibility that the FF is an imperfect financial indicator, the authors extend the analysis to consider the forecasting power of individual financial variables. They find their result about lack of predictability of financial variables to be largely robust. One important exception, which I return to later, is that credit spreads provide helpful information for near term mean of GDP growth, consistent with results elsewhere in the literature.

TWO CLARIFICATIONS There are two aspects of the analysis that are helpful to clarify. First, the emphasis on the marginal information from financial variables is one important way the paper differs from earlier literature (and could account for some of the differences in findings). In contrast to Adrian, Boyarchenko, and Giannone (2019), the FF excludes the information from the contemporaneous interaction between the financial and real sectors (contained in the GF). By doing so, the authors isolate information from financial conditions that is purely forward-looking. Given the objective of designing an early warning system, that is, a financial siren about risks to future growth that could go off independently of current economic conditions, the authors' approach makes sense. However, as the authors clearly recognize, it is important to keep in mind that their forecasting exercise is silent on the importance of financial conditions for economic activity. Most of the theories of financial–real sector interactions they cite are based on contemporaneous mutual feedback, information that is excluded in their forecasting exercise.

Second, the use of an index for financial conditions that aggregates both credit prices and credit quantities is problematic. Credit prices and quantities differ in cyclical behavior. In particular, credit aggregates tend to oscillate at lower frequencies than spreads. As I show below, they also have a longer lead over real activity than do spreads. In addition, the economic interpretation of quantity and spread behavior can differ, as I also discuss below. In the end, aggregating credit quantity and price information makes the results difficult to interpret. The authors recognize this issue and, as I noted earlier, address it by also considering individual financial variables.

SOME PICTURES TO TELL THE STORY Figures 1 and 2 help illustrate the issues underlying the authors' findings. Figure 1 portrays some basic features of a financial crisis based on evidence from Krishnamurthy and Muir (2017). The data are annual from a panel of industrialized economies from 1869 to 1918. Each panel plots the average behavior of a variable before, during, and after the crisis. The upper right-hand panel shows that roughly three

Figure 1. Behavior of Credit Spreads, GDP, and the Quantity of Credit

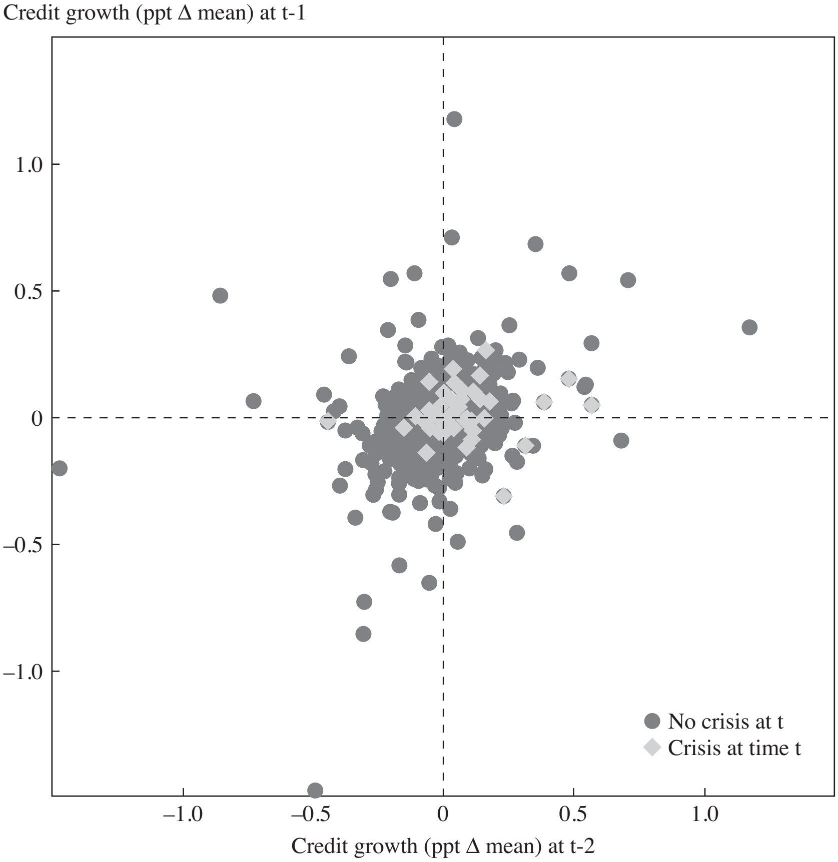


Source: Krishnamurthy and Muir (2017).

Note: This figure plots the behavior of credit spreads, GDP, and the quantity of credit around a financial crisis with the crisis beginning at time zero. GDP and credit are expressed in deviation from (country specific) trend. Spreads are normalized by dividing by the unconditional mean.

years prior to a crisis a credit boom emerges (typically associated with increasing asset prices). The upper left-hand panel shows that roughly two years later, on the eve of the crisis, credit spreads increase steadily and peak just following the crisis. From the lower panel, we see that there is a mild output boom entering the crisis followed by a sharp contraction. The figure clearly reveals nonlinear behavior of output in a financial crisis. It also shows the distinct patterns of credit aggregates and credit spreads. The former tend to exhibit a longer lead over the crisis than the latter. A natural interpretation is as follows. The buildup of credit (and leverage) increases borrowers' vulnerability to negative shocks. When negative shocks (e.g., declining asset prices) eventually occur, borrower balance sheets weaken and financial distress emerges. Credit spreads reflect this distress. The important point to note is that while credit quantity and spreads play interrelated roles in the crisis, their timing and economic

Figure 2. Credit Booms and Financial Crises



Source: Gertler, Kiyotaki, and Prestipino (forthcoming).
Note: Run Frequency after boom: 4.9 pct.; After no boom: 2.8 pct.; (Boom: top right quadrant).

relevance is distinct. For this reason, financial indexes that aggregate both types of variables may not provide the most efficient use of information.

Figure 2, taken from Gertler, Kiyotaki, and Prestipino (forthcoming), illustrates why it may be difficult to find a reliable early warning signal of a financial crisis, despite the evidence in figure 1 that credit growth tends to lead crises by at least several years. The central identification problem is that there are "good" credit booms as well as "bad" ones, with the former being far more prevalent than the latter. The horizontal axis in figure 2 is demeaned credit growth for a country two years prior to current time. The vertical axis is demeaned credit growth one year prior. We can then define a credit boom as two years of above average growth, which corresponds to the upper right-hand quadrant in the figure. The diamond-shaped dots in the figure are times within a country where a crisis occurred. while the round dots are times absent a crisis. The key message is that conditional on a credit boom (i.e., conditional on being in the upper right-hand quadrant) crises occur only 5 percent of the time. That is, most of the time credit booms are good. It is true that a crisis is more likely conditional on boom. Conditional on no boom a crisis occurs with only 2.8 percent probability. The bottom line, however, is that credit growth is unlikely to provide a reliable early warning signal.

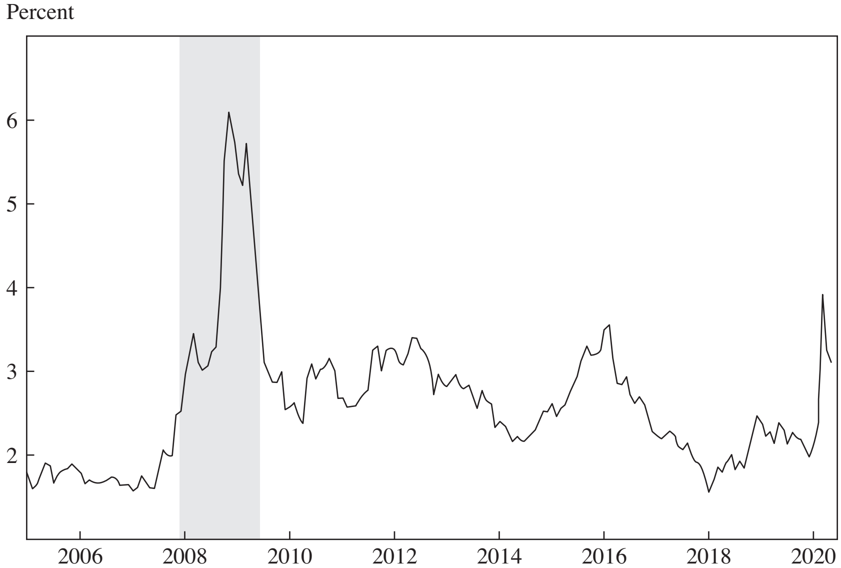
Credit spreads are likely a more reliable indicator than credit aggregates. However, as figure 1 shows, the lead time of a spike in spreads over a crisis is typically much shorter than that of credit growth. Thus, spreads are unlikely to provide a lengthy advance warning of the crisis. Nonetheless, the authors do present evidence that, for the near term, spreads help forecast the mean of GDP growth, which is consistent with results elsewhere in the literature.

One caveat the authors note is that spreads do not provide information about the higher moments of GDP growth. Nonetheless, spreads may provide information about the depth of a recession. The reason is that spreads exhibit asymmetric positive jumps on the eve of crises that mirror the asymmetric declines in output, as figure 1 makes clear. A large jump in the spread prior to a crisis, accordingly, predicts a large drop in the conditional mean of output growth, everything else being equal.

A FEW THOUGHTS GOING FORWARD As the paper makes clear, a central challenge in forecasting the distribution of GDP growth involves data limitations. There are simply not enough data to get statistical precision in the link between financial indicators and the higher moments of GDP growth. In the case of the United States, there has been only one major financial crisis in the postwar period (though that may change depending on how the current crisis plays out). To be sure, financial stress has occurred in previous downturns, but not on the same scale. One possibility, which the authors pursue at the end of the paper, is to exploit international data. The advantage, of course, is that they will gain more observations on financial crises. Here they find only very limited success: credit spreads are helpful in forecasting mean GDP growth (as noted earlier), but otherwise financial variables have little information content about GDP. As the authors note, though, more needs to be done to account for country heterogeneity.

It might also be useful to consider alternative financial indicators. For example, Krishnamurthy and Muir (2017) find success by interacting credit

Figure 3. Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity



Source: Federal Reserve Economic Data (FRED).

aggregates with credit spreads. Another interesting possibility might be to disentangle unusual periods of high credit growth from normal periods of robust growth, as done by Hasenzagl, Reichlin, and Ricco (2020). Here, though, it might be necessary to make use of international data, given the limited number of such episodes in the United States.

BRIEF OBSERVATIONS ON FINANCIAL INDICATORS IN THE CURRENT CRISIS By mid-March, when it was obvious that the pandemic was going to have a clear economic effect, credit spreads increased sharply. The Baa corporate bond spread, as portrayed in figure 3, rose sharply from 200 basis points to nearly 425 basis points. This was below the peak of the Great Recession but at the same level at similar stages of the recession. Other financial indicators also pointed to distress, including plunging stock prices, various measures of liquidity in bond markets, and an increasing VIX. In this regard, financial variables were incorporating news of the contraction that would follow.

Were it not for an aggressive intervention in financial markets, financial market conditions would likely have continued to deteriorate sharply. Among other things, the Federal Reserve committed to buying high-grade corporate bonds with the backing of the Treasury. In the period following the announcement, corporate spreads have declined roughly 100 basis points and other financial distress measures have receded as well. For our purposes, the key lesson is that by (appropriately) taking action to reduce spreads the Federal Reserve has likely reduced the information content of this variable.

CONCLUDING REMARKS This paper makes a convincing case that the marginal information that financial variables have for the distribution of GDP growth is minimal. It is well executed and a very useful contribution to the forecasting literature.

As the authors would agree, though, lack of forecasting power is not the same thing as lack of importance for real activity. Much of the real–financial interaction is contained in the contemporaneous interaction between real and financial variables, which is excluded from the authors' definition of marginal financial information.

The key lesson for policy is that macroprudential policy should not be based on predictability. Rather, it should design the best response to unpredictable shocks that disrupt the financial system. This design, further, will most certainly depend on financial variables (e.g., bank leverage ratios, liquidity measures, etc.).

REFERENCES FOR THE GERTLER COMMENT

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone. 2019. "Vulnerable Growth." *American Economic Review* 109, no. 4: 1263–89.

Gertler, Mark, Nobuhiro Kiyotaki, and Andrea Prestipino. Forthcoming. "Credit Booms, Financial Crises and Macroprudential Policy." *Review of Economics Dynamics*.

Hasenzagl, Thomas, Lucrezia Reichlin, and Giovanni Ricco. 2020. "Financial Variables as Predictors of Real Growth Vulnerability." Discussion Paper 14322. London: Centre for Economic Policy Research. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=14322.

Krishnamurthy, Arvind, and Tyler Muir. 2017. "How Credit Cycles across a Financial Crisis." Working Paper 23850. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w23850.

COMMENT BY

NELLIE LIANG This paper tackles two broad questions: Do financial variables have predictive value for GDP growth, and can higher moments of the GDP growth distribution be predicted? It is a forecasting paper, and it is extensive. The authors argue that financial variables have no predictive power for GDP growth or risk to growth. But the paper misses the broader and more important relationship between the financial sector and real economy, which I discuss below and illustrate with the example of recent Federal Reserve actions in response to COVID-19.

I agree with the authors' point that financial variables cannot be reliable predictors of a crisis, like one brought on by a pandemic or even the 2008 global financial crisis. A crisis represents the product of a negative shock and vulnerabilities, which are amplifiers of shocks. To predict a crisis, financial variables would need to be able to predict both a negative shock and financial vulnerabilities. And they are not better at predicting a shock than any other variables.

But the message of the paper should not be that financial variables are not useful indicators of vulnerabilities in the financial sector because they can't predict shocks. It would be a mistake to suggest that policymakers can ignore the effects of financial variables on the economy.

Financial stability reports of central banks are designed to guide macroprudential policies and to be inputs into monetary policy and emergency liquidity actions. The reports are very careful to say they cannot predict the next crisis, but they highlight the conditions that make the economy more prone to a crisis or deep recession because financial vulnerabilities are high and could amplify any unexpected negative shock through fire sales or contagion. Macroprudential policies aim to preemptively reduce financial vulnerabilities, like raising capital requirements for banks when credit is booming so that banks will have additional capital buffers to absorb higher future losses and be able to continue to provide credit should there be a negative shock. Indicators of financial vulnerabilities are financial variables, like asset valuations, credit burdens of borrowers, and leverage and funding risks of financial intermediaries. But in this paper the authors argue that financial variables have no predictive power and convey a message based on that narrow framing that they have no value for thinking about economic risks.

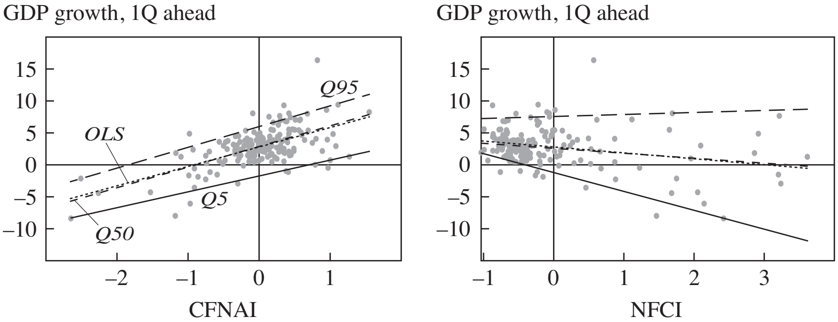
In its May 2020 financial stability report, the Federal Reserve highlights how some hedge funds and private mortgage funds quickly sold assets after asset prices fell in response to the spread of COVID-19. These sales contributed to unusual market dysfunction in the US Treasury and mortgage-backed securities markets as bid-ask spreads widened and market depth shrank. Had financial regulators been willing or able to take more forceful macroprudential actions in recent years to limit the systemic consequences of leverage, liquidity risk, and model-driven strategies of some private funds, the fall in asset prices brought on by the pandemic (which could not be predicted) would not have been amplified to such an extent and could have limited purchases of Treasury securities and residential and commercial mortgage-backed securities by the Federal Reserve in March. (Those purchases were emergency liquidity actions, not quantitative easing or macroprudential actions.) The point is that financial vulnerabilities can have significant consequences for growth and risks to growth.

The authors argue that financial variables have no predictive power for GDP growth or risk to growth because "markets do not anticipate the timing of [a] recession and they price the risk only once they see it." This statement highlights the framing of their empirical analysis, that financial variables are valuable only if they can predict shocks that can lead to recessions before real data can predict them. For several reasons and illustrated above with a specific example, I think the framing in the paper is too narrow. First, it ignores a more relevant question for financial policymaking of whether financial variables affect risk-taking behavior of borrowers and lenders which lead to financial vulnerabilities. The importance of behavioral effects, indeed the endogeneity of financial variables and real activity, is a primary lesson of the global financial crisis and is core to macroprudential policymaking. That is, financial conditions can affect the buildup of financial vulnerabilities which can amplify large negative shocks, such as the recent COVID-19 pandemic. These amplification effects show through as financial market dysfunction and restricted credit supply which can separately increase risks to growth.

A second reason is that here the authors use a broad activity indicator that includes both real and financial variables, as well as price and monetary policy variables, and then evaluates whether a separate financial variable constructed to be orthogonal to the broad indicator has additional predictive value. This unique construction leads to a financial factor that differs from most others and may be biased against finding value in financial variables leading to findings that seem to differ from those in a growing list of other studies. Third, when evaluating whether individual indicators might be significant rather than indexes, the authors do not distinguish between types of financial variables. That is, the tests are horse races of variables without imposing any structure, which the authors acknowledge. But many papers find that the effects of different variables differ significantly, because funding risks will be different from the effects of credit risk and from the effects of financial intermediary risks.

In my discussion below, I focus on the results and interpretation related to forecasting the mean and variance of the US real GDP growth distribution

Figure 1. Quantile Regressions of GDP Growth on Real Activity and Financial Conditions Indexes



Source: Author's calculations.

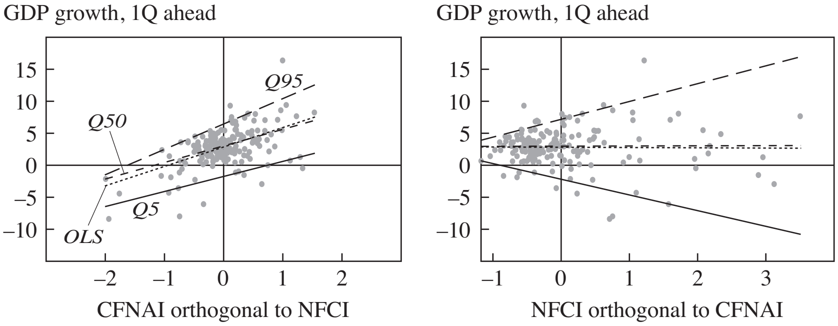
and interpret the results in terms of implications for macroprudential policy. The empirical results on skewness and kurtosis, as well as forecasts for other countries, do not change their broad conclusions.

To illustrate the issue of whether financial variables offer any additional information above the content in real activity variables, I estimate quantile regressions for GDP growth one quarter ahead on two indicators, an index of real economic activity (Chicago Fed National Activity Index, CFNAI) and a separate financial conditions index (NFCI). Both indexes are constructed by the Federal Reserve Bank of Chicago and allow the estimation period to start in 1975. In figure 1, I show the scatterplots of GDP growth one quarter ahead plotted, first, against the index of real economic activity and, second, against the financial conditions index. The lines in the plots are slopes of the quantile regressions. In contrast to ordinary least squares, quantile regressions minimize the absolute deviation rather than squared deviation of errors; this has the effect of weighting errors more heavily near the quantile of interest than errors that are further away.

As shown, the median quantile for GDP growth one quarter ahead for the real activity index (CFNAI) has a positive slope, and the slopes for the fifth percentile and the ninety-fifth percentile are similar, suggesting the variance is constant across different activity levels.

For the financial conditions index (NFCI), the slope of the median is slightly negative (tighter financial conditions in the current quarter, lower GDP growth in the next quarter). In contrast to the results for CFNAI, the slope of the fifth percentile is much more negative than for the median.

Figure 2. Quantile Regressions of GDP Growth on Orthogonal Real and Financial Conditions Indexes



Source: Author's calculations.

These slopes are statistically different. They show the variance differs, although there clearly are fewer data points at higher levels of the NFCI, when they are tighter and GDP growth one quarter ahead is lower.

The next two charts in figure 2 illustrate the same relationships but are based on indexes now constructed so they are orthogonal to each other. The quantile regression slopes of the real activity index that is purged of the financial index and GDP growth one quarter ahead does not change. The slopes of the NFCI purged of the real index to GDP growth one quarter ahead also do not change much. The fifth percentile remains more negatively related to financial conditions than the median, and the coefficients are statistically different.

These charts are a straightforward way to illustrate that financial variables do matter for variance once real activity variables are included. This real activity index does not include financial variables, and so a financial indicator orthogonal to the real index still has predictive value. In contrast, in this paper the authors use a global factor, which is common to all 112 variables in the FRED-MD data set, which are real, price, monetary, and financial. The financial factor is constructed from the financial variables but also orthogonal to the global factor. A financial factor is then interpreted to be important only if it is important separately and in addition to its role in the global factor.

A number of studies other than Giglio, Kelly, and Pruitt (2016) and Adrian, Boyarchenko, and Giannone (2019), which the authors cite in the paper, have found significant effects of various financial variables on GDP risk in the United States. For example, Coe and Vahey (2020) use non-Gaussian and nonlinear estimations to predict risk to growth in the four crisis periods since 1875; Kiley (2018) uses quantile regressions to predict risks to the unemployment rate in the United States; Carriero, Clark, and Marcellino (2020a) and Caldara, Scotti, and Zhong (2020) use vector autoregression models with stochastic volatility to capture tail risks; and Carriero, Clark, and Marcellino (2020b) test a number of alternative models and show that a number of financial variables improve both point and tail risk nowcasts of GDP. Financial variables have been found to be significant for risk in other countries as well, in work by Chavleishvili and Manganelli (2019) for the euro area and Duprey and Ueberfeldt (2020) for Canada.

The authors also evaluate the time-varying distribution of GDP growth and find that parameter uncertainty around time-varying moments other than the mean are imprecisely estimated. While the standard deviation of GDP growth one quarter ahead clearly varies over time, and in line with the mean, the paper shows it has high parameter uncertainty around recession periods. The authors emphasize that the variance cannot be estimated precisely and the financial factor is not important. I would emphasize instead the finding that the GDP growth distribution is time-varying and that periods when uncertainty is higher—the recessions—are precisely the periods we care most about. Rather than concluding it is a futile exercise to estimate these periods, I think a more appropriate interpretation is that more work is needed to help predict these important events.

One extension could be to draw on research that distinguishes financial indicators by specific concepts they are intended to measure. To illustrate, Bernanke (2018) uses daily data on seventy-five financial variables and aggregates them to a monthly frequency to evaluate the effects of financial variables on mean GDP growth. Importantly, he splits the data to represent four groups, reflecting housing and mortgages, nonmortgage credit availability, short-term funding, and bank solvency. He finds the effects of the four factors vary significantly: panic factors (credit and funding) are significant predictors of the means of monthly GDP, industrial production, the unemployment rate, and other variables, whereas the balance sheet factors are less significant. He does not test for variance, but the sample period contains only one recession.

Many papers that have studied the role of credit on large output losses in the future have incorporated credit cycles of many years because cycles can take a while to emerge. For example, Jordà, Schularick, and Taylor

(2013) focus on credit growth since the trough of the last recession, and Kiley (2018), Aikman and others (2020), and Adrian and others (2018) use between eight and sixteen quarters to capture credit cycles. In a separate paper, Hasenzagl, Reichlin, and Ricco (2020) show that the leverage subcomponent can help to predict the variance of quarterly GDP growth, while the subcomponents for market and credit risk do not. This type of analysis is missing in this paper.

In a third stage of the analysis, the authors test the predictive value of individual financial variables for the distribution of GDP growth. They use data in FRED-QD and add some additional data. The data set for the United States has forty-three variables, of which fifteen are financial variables, a mix of quantity and price variables. The authors employ a conditional heteroskedasticity model for GDP growth and a method that selects the relevant predictor variables among a large set of variables using a "Bayesian prior distribution on the model parameters that imposes approximate sparsity." Each of the variables are entered separately. They find that only a couple of financial variables can help to predict mean or variance, though perhaps this is not surprising given many of the variables are collinear and they are not grouped in any way.

What does this paper suggest for macroprudential policy? Macroprudential policies are designed to increase the resilience of the financial sector to negative shocks, to reduce amplification because disruptions in credit and funding can have serious repercussions for the real economy. Financial variables can help macroprudential policymakers for what they signal about possible buildups of financial vulnerabilities when financial conditions allow lenders and borrowers to increase risk taking by more than usual, which would make the financial system more vulnerable to negative shocks. Financial variables can also signal possible disruptions in market liquidity and credit supply which if sustained would increase risks to growth as borrowers lost access to credit.

The authors acknowledge that there is a role for macroprudential policies because financial vulnerabilities can amplify shocks. But their results aren't directly applicable because they test the value of financial variables mainly as noisy predictors that are just a reflection of real activity, rather than an assessment of current and potential buildup of financial imbalances.

REFERENCES FOR THE LIANG COMMENT

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone. 2019. "Vulnerable Growth." *American Economic Review* 109, no. 4: 1263–89.

Adrian, Tobias, Federico Grinberg, Nellie Liang, and Sheheryar Malik. 2018. "The Term Structure of Growth-at-Risk." Working Paper 42. Washington: Hutchins Center on Fiscal and Monetary Policy, Brookings Institution. https://www.brookings.edu/wp-content/uploads/2018/08/WP42-NL-2.7.pdf.

Aikman, David, Andreas Lehnert, Nellie Liang, and Michele Modugno. 2020. "Credit, Financial Conditions, and Monetary Policy Transmission." *International Journal of Central Banking* 16, no. 3: 141–79.

Bernanke, Ben S. 2018. "The Real Effects of the Financial Crisis." *Brookings Papers on Economic Activity*, Fall, 251–342.

Caldara, Dario, Chiara Scotti, and Molin Zhong. 2020. "Macroeconomic and Financial Risks: A Tale of Volatility." Working Paper.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino. 2020a. "Capturing Macroeconomic Tail Risks with Bayesian Vector Autoregressions." Working Paper 20-02. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202002.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino. 2020b. "Nowcasting Tail Risks to Economic Activity with Many Indicators." Working Paper 20-13. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202013.

Chavleishvili, Sulkhan, and Simone Manganelli. 2019. "Forecasting and Stress Testing with Quantile Vector Autoregression." Working Paper 2330. European Central Bank. https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2330~85709114d4.en.pdf.

Coe, Patrick J., and Shaun P. Vahey. 2020. "Financial Conditions and the Risks to Economic Growth in the United States since 1875." Working Paper 36/2020. Canberra: Crawford School of Public Policy, Australian National University. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3582337.

Duprey, Thibaut, and Alexander Ueberfeldt. 2020. "Managing GDP Tail Risk." Working Paper 20-3. Bank of Canada. https://www.bankofcanada.ca/wp-content/uploads/2020/01/swp2020-3.pdf.

Federal Reserve Board of Governors. 2020. *Financial Stability Report*, May. https://www.federalreserve.gov/publications/2020-may-financial-stability-report-purpose.htm.

Giglio, Stefano, Bryan Kelly, and Seth Pruitt. 2016. "Systemic Risk and the Macroeconomy: An Empirical Evaluation." *Journal of Financial Economics* 119, no. 3: 457–71.

Hasenzagl, Thomas, Lucrezia Reichlin, and Giovanni Ricco. 2020. "Financial Variables as Predictors of Real Growth Vulnerability." Discussion Paper 14322. London: Centre for Economic Policy Research. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=14322.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2013. "When Credit Bites Back." *Journal of Money, Credit and Banking* 45, no. s2: 3–28.

Kiley, Michael T. 2018. "Unemployment Risk." Finance and Economics Discussion Series 2018-067. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/econres/feds/files/2018067pap.pdf.

GENERAL DISCUSSION James Stock thanked the authors and discussants for a great paper and informative discussion. He said he was sympathetic to the challenge of having to make forecasts of not just GDP but also its moments using financial variables. He wondered why the authors' results differed so much from those of Adrian, Boyarchenko, and Giannone.[1] He asked whether the distinctions were due to differences in methods or indexes.

Lucrezia Reichlin pointed out she had recently used exactly the same data as Adrian, Boyarchenko, and Giannone to test the performance of financial variables in out-of-sample forecasting in a paper together with Ricco and Hasenzagl.[2] This paper found similarly negative results about the ability to forecast GDP using financial variables, so the difference could not be the data. Instead, the authors' results differ substantially from those of Adrian, Boyarchenko, and Giannone because of how the global and financial indexes are constructed, she said. Several financial indexes, like the Chicago Fed's National Financial Conditions Index, are actually very correlated with real factors. This fact motivated the authors to construct a financial index that separates out all of the effect of real variables to answer the question: How much additional predictive power does the purely financial component of the movement in financial variables give you? As it turns out, the answer is not much.

Nevertheless, Reichlin agreed with Liang's point that financial variables are very important for the business cycle. But much of that effect works endogenously through common factors captured by the authors' global factors. Reichlin noted that the authors take a "brain-dead forecasters" approach, ignoring any structural mechanism relating financial and real variables, and focus purely on the usefulness of their financial factors for forecasting GDP and its moments. They would need to take a different, more structural, approach to get at the endogenous mechanisms driving the effect of financial variables on business cycles, Reichlin said.

Giovanni Ricco noted that both Liang and Gertler questioned whether it was possible to cleanly separate real and financial shocks in their discussions. Ricco responded that their paper does not perform a fully structural exercise, and a structural model is needed to get at the mechanisms through

1. Tobias Adrian, Nina Boyarchenko, and Domenico Giannone, "Vulnerable Growth," *American Economic Review* 109, no. 4 (2019): 1263–89.

2. Lucrezia Reichlin, Giovanni Ricco, and Thomas Hasenzagl, "Financial Variables as Predictors of Real Growth Vulnerability," Discussion Paper 05/2020 (Frankfurt: Deutsche Bundesbank, 2020), https://www.bundesbank.de/resource/blob/827682/3deb1560a27f63fe08d2f60628eb7636/mL/2020-03-05-dkp-05-data.pdf.

which financial and real variables interact. Having said that, the starting point of their paper was that movements in financial variables reflect some mixture of information about developments in the real economy and extra financial stresses. This motivated them to take the admittedly extreme approach of removing all of the global and real factors and to instead work with just the orthogonal financial component. This decision allowed them to analyze whether the financial component had additional predictive power. If so, this finding would provide strong reduced form evidence in favor of an additional financial frictions story. However, their results indicate that their index does not have much predictive power. Ricco noted that this result is not the end of the story. He noted that extending the analysis with more structural guidance is necessary for thinking about policy.

In response to Liang's point that estimates of the standard deviation of GDP do move around over time, Ricco pointed out that while it is true that their point estimates move around, it is still possible to draw a flat straight line through their uncertainty band. While their paper shows that it is certainly true that financial variables have some predictive power for the variance of GDP, Ricco observed that overall the results on the second moment showed a very weak predictive relation.

DARON ACEMOGLU
Massachusetts Institute of Technology

ANDREA MANERA
Massachusetts Institute of Technology

PASCUAL RESTREPO
Boston University

Does the US Tax Code Favor Automation?

ABSTRACT We argue that the US tax system is biased against labor and in favor of capital and has become more so in recent years. As a consequence, it has promoted levels of automation beyond what is socially desirable. Moving from the US tax system in the 2010s to optimal taxation of capital and labor would raise employment by 4.02 percent and the labor share by 0.78 percentage point and restore the optimal level of automation. If moving to optimal taxes is infeasible, more modest reforms can still increase employment by 1.14–1.96 percent, but in this case it is also beneficial to impose an additional automation tax to reduce the equilibrium level of automation. This is because marginal automated tasks do not bring much productivity gains but displace workers, reducing employment below its socially optimal level. We additionally show that reducing labor taxes or combining lower capital taxes with automation taxes can increase employment much more than the uniform reductions in capital taxes enacted between 2000 and 2018.

Conflict of Interest Disclosure: Daron Acemoglu is Institute Professor at the Massachusetts Institute of Technology (MIT) and a member of the Temporary Nomination Committee for the National Academy of Sciences; Andrea Manera is a graduate student in economics at MIT; Pascual Restrepo is an assistant professor of economics at Boston University and recently served as guest editor for the Journal of Labor Economics. Beyond these affiliations, the authors did not receive financial support from any firm or person for this paper or from any firm or person with a financial or political interest in this paper. They are currently not officers, directors, or board members of any organization with an interest in this paper. No outside party had the right to review this paper before circulation. The views expressed in this paper are those of the authors, and do not necessarily reflect those of MIT, Boston University, or the National Academy of Sciences.

The last three decades have witnessed a declining share of labor in national income, stagnant median real wages, and lower real wages for low-skill workers in the US economy (Elsby, Hobijn, and Şahin 2013; Acemoglu and Autor 2011; Karabarbounis and Neiman 2014). The labor share in nonfarm private businesses declined from 63 percent in 1980 to 56 percent in 2017, while median real wages grew only by 16 percent (as compared to GDP per capita which doubled during the same period), and the real wages of male workers with a high school diploma fell by 6 percent between 1980 and 2017. In the meantime, production processes have become increasingly automated, as computerized numerical control machines, industrial robotics, specialized software, and, lately, artificial intelligence technologies have spread rapidly throughout the economy. For instance, the US economy had a total of 2.5 industrial robots per thousand workers in manufacturing in 1993, and this number rose to 20 by 2019 (Acemoglu and Restrepo 2020). From a base of essentially zero in the mid-2000s, the share of vacancies posted for artificial intelligence–related activities increased to 0.75 percent by 2018 (Acemoglu and others 2020).

A common perspective among economists is that even if automation is contributing to the decline in the labor share and the stagnation of wages, the adoption of these technologies is beneficial, and any adverse consequences should be dealt with using redistributive policies and investments in education and training. But could it be that the extent of automation is excessive, meaning that businesses are adopting automation technologies beyond the socially optimal level? If this is the case, the policy responses to this trend need to be rethought.

In this paper, we show that the US tax system is biased against labor and as a result generates excessive automation and suboptimally low levels of employment and labor share. We first introduce a task-based model of automation, building on Acemoglu and Restrepo (2018, 2019a, 2019b) and Zeira (1998), to study the interplay between taxes and automation. Our first theoretical result establishes that optimal capital and labor taxes depend on the inverse supply elasticities of these factors and labor market frictions. Consistent with Diamond and Mirrlees (1971), once capital and labor taxes are set optimally, there is no reason to distort equilibrium automation decisions. Intuitively, optimal taxes undo any distortions and ensure that market prices reflect the social values of capital and labor. Automation decisions based on these prices are therefore optimal.[1]

1. We assume that the labor market friction is common across tasks. When labor market frictions affect tasks differentially, there is an additional reason for excessive automation,

Yet this result does not imply that equilibrium automation decisions are optimal at arbitrary capital and labor taxes. Our second theoretical result shows that if a tax system is biased against labor and in favor of capital—that is, taxes on labor are too high and taxes on capital are too low—then reducing automation at the margin improves welfare. We show that this reduction can be achieved with an *automation tax*, which is an additional tax on the use of capital in tasks where labor has a comparative advantage. An automation tax is beneficial because reducing automation below its equilibrium level has second-order costs and first-order benefits. The costs are second-order as the productivity gains from automating marginal tasks are small, or, equivalently, the automation of marginal tasks corresponds to "so-so automation" in the terminology of Acemoglu and Restrepo (2019a, 2019b). But when the tax system is biased against labor and thus the level of employment is below the social optimum, limiting automation and avoiding the resulting displacement of labor has first-order benefits.

A common intuition is that if taxes are distorted, then the best policy remedy is to correct these distortions. Hence, if a tax system treats capital too favorably, we should directly tackle this distortion and increase capital taxes. We demonstrate that this intuition does not always apply in the presence of other constraints—for example, a lower bound on labor taxes. Our third theoretical result shows that a tax system distorted in favor of capital may call for reducing equilibrium automation even if raising capital taxes is possible. In fact, when moving to the unconstrained optimum is not feasible, constrained optimal policy may involve lower capital taxes in addition to a reduced level of automation because this combination avoids the displacement of workers from marginal tasks while ensuring that capital gets used intensively in tasks that are (and should be) automated. Both of these margins contribute to raising employment and welfare. This result underscores the importance of distinguishing between the choice of capital intensity in tasks where capital has a comparative advantage and automation, which involves the substitution of capital for labor in additional tasks. An automation tax is beneficial precisely because it does not reduce capital intensity uniformly but discourages the automation of marginal tasks.

Armed with these theoretical results, we turn to measuring effective taxes on capital and labor in the United States and comparing them to their optimal counterparts. We find that labor is much more heavily taxed than capital, and this difference has increased in recent years. Effective labor

as shown in our companion paper (Acemoglu, Manera, and Restrepo in progress), and in that case, distorting automation may be beneficial even when taxes are set optimally.

taxes in the United States are in the range of 25.5–33.5 percent. Effective capital taxes on software and equipment, on the other hand, are much lower, 10 percent in the 2010s and 5 percent after the 2017 tax reforms, though they used to be about 20 percent in 2000.[2] About half of this decline is due to the greater generosity of depreciation allowances.

Using plausible ranges for the elasticities of the capital and labor supply and estimates of labor market distortions, we find that the US tax system is biased against labor. In fact, our baseline estimates suggest that optimal labor taxes are lower than capital taxes—an 18.22 percent labor tax compared to a 26.65 percent capital tax. Optimal taxes are lower for labor than for capital because empirically plausible ranges of supply elasticities for capital and labor are similar, but employment is further distorted by labor market imperfections. Moving from the current tax system to optimal taxes would reduce the range of automated tasks by 4.1 percent and increase employment by 4.02 percent and the labor share by 0.78 percentage point.[3]

Our quantitative results show that, as in our theory, reducing automation is socially beneficial. Specifically, with no changes in capital and labor taxes, an automation tax of 10.15 percent—which implies that only tasks where the substitution of labor for capital reduces unit costs by more than 10.15 percent are automated—maximizes welfare and raises employment by 1.14 percent and the labor share by 1.93 percentage points. If capital taxes can be reduced as well, then a 12.9 percent automation tax combined with a reduction in capital taxes from 10 percent to 8.39 percent would achieve even higher welfare gains and increase employment by 1.59 percent and the labor share by 2.44 percentage points. We further show that tax reforms that involve lower labor taxes or combine lower capital taxes with an automation tax would have increased welfare and expanded employment much more than the uniform capital tax reductions enacted between 2000 and 2018.

We conclude with two extensions. First, we show that if human capital is endogenous, the asymmetric treatment of labor becomes more costly as it distorts human capital investments, leading to even lower optimal taxes on labor and more excessive automation under the current system. Second, we consider endogenous development of automation technologies,

2. Acemoglu and Restrepo (2019b) document that technological changes in the four decades after World War II involved less automation and more rapid advances in technologies that increased human productivity (such as the creation of new tasks for workers) than has been the case recently. Though there are other reasons for why the direction of technology altered, the lower taxation of equipment and software capital may have also played a role.

3. Despite these large changes in employment, the increase in welfare is given by a Harberger's triangle and is thus smaller—0.38 percent in consumption-equivalent terms.

which come at the expense of other types of innovations that are more beneficial for labor. In this case, there are reasons for not just preventing excessive adoption of automation technologies but also redirecting technological change away from further automation (and this is true even with optimal taxes on capital and labor).

Our paper is related to several classic and recent articles, though, to the best of our knowledge, no other paper investigates whether the US tax system favors automation.

First, there is an emerging literature on redistribution and taxation of automation technologies (Guerreiro, Rebelo, and Teles 2017; Thuemmel 2018; Costinot and Werning 2018). This literature studies whether adverse distributional effects of automation call for taxes on automation technologies. Our paper is complementary to this literature, as it focuses on situations in which the tax system is biased against labor and the key policy objective is to raise employment (not to redistribute income).

Second, our paper is related to the literature on optimal capital taxation (e.g., Atkinson and Stiglitz 1972, 1976; Judd 1985; Chamley 1986; Straub and Werning 2020). Our contribution is to show that in both two-period and infinite-horizon settings, provided that the government must run a balanced budget at each date, optimal taxes are given by the same inverse-elasticity formulas (with an additional term adjusting for labor market frictions). In contrast, this literature typically assumes that the government can freely accumulate assets and concludes that zero capital taxation is optimal in the long run. Straub and Werning (2020) show that if the supply of capital is not perfectly elastic (which means utility is not time-additive), then the government accumulates sufficient assets so that both capital and labor face zero taxes in the long run. We demonstrate in the online appendix that in the presence of labor market frictions, the same reasoning leads to a subsidy to labor. Thus, in the empirically relevant case of a finite supply elasticity of capital, even without the balanced budget assumption, the US tax system with low capital taxes and high labor taxes is far from optimal.

Third, our paper relates to the literature on the effects of tax reforms on investment and labor market outcomes. A branch of this literature estimates the differential responses of investment across firms facing different taxes (Goolsbee 1998; Hassett and Hubbard 2002; Edgerton 2010; Yagan 2015).[4]

4. Modal results in this literature find investment elasticities with respect to the keep rate (one minus the tax rate) between 0.5 and 1. More recent work by House and Shapiro (2008) documents a larger investment response and argues that this was due to the temporary nature of the bonus, while Zwick and Mahon (2017) estimate investment elasticities with respect to the keep rate that are around 1.5 for most firms.

However, these estimates are informative about firms' demand for capital, not about the (long-run) elasticity of the supply of capital, which is the relevant object for optimal taxes. We discuss below estimates of this elasticity based on the response of the supply of capital to wealth and capital income taxes (see Kleven and Schultz 2014; Zoutman 2018; Brülhart and others 2016; Jakobsen and others 2020; Durán-Cabré, Esteller-Moré, and Mas-Montserrat 2019). More closely connected to our work is a branch of this literature on the labor market implications of tax reforms. Suárez Serrato and Zidar (2016) exploit the incidence of tax changes across US counties and estimate that a 1 percent increase in the keep rate of corporate taxes raises employment by 3.5 percent and wages by 0.8 percent and that workers bear 35 percent of the incidence. Garrett, Ohrn, and Suárez Serrato (2020) compare counties at the 75th percentile of exposure to bonus depreciation allowances to those at the 25th percentile and find a 2 percent increase in employment, no changes in wages, and a 3.3 percent increase in investment in response to the reform. These estimates point to a fairly elastic response of employment and a less than perfectly elastic response of capital in local labor markets (a perfectly elastic response of capital would cause workers to bear the full incidence).

Finally, our modeling of automation builds on Zeira (1998), Autor, Levy, and Murnane (2003), Acemoglu and Autor (2011), and most closely, Acemoglu and Restrepo (2018, 2019a, 2019b). The task-based framework is useful in our setting because it shows how automation (substituting capital for labor in tasks previously performed by humans) creates a displacement effect while automating marginal tasks generates limited productivity gains (because firms are approximately indifferent between automating these tasks or producing with labor). This combination of displacement effects and small productivity gains is at the root of our result that the planner would like to reduce automation at the margin when the tax system is biased against labor. Our framework also clarifies how policy can affect the level of automation and why taxing automation is not the same as taxing capital.

The rest of the paper is organized as follows. Section I introduces our conceptual framework and derives our theoretical results. Section II provides a detailed discussion of the US tax system and maps the complex US tax code into effective capital and labor income taxes. Section III then explores whether these taxes are biased and how they compare against optimal taxes. Section IV discusses two extensions of our framework, while section V concludes. The online appendix contains proofs of the results stated in the text, various theoretical generalizations, and further details for and robustness checks on our empirical work.

I. Conceptual Framework

This section presents our conceptual framework for evaluating the optimality of capital and labor taxes and the extent of automation. To facilitate the exposition, we focus on a two-period model and generalize our main results to an infinite-horizon setting in the online appendix.

I.A. Environment

There is a unique final good, produced at time t = 1 by combining a unit measure of tasks:

$$\left(\int_0^1 y(x)^{\frac{\lambda-1}{\lambda}} dx\right)^{\frac{\lambda}{\lambda-1}}.$$

Tasks are allocated between capital and labor and performed with the following task-level production function:

$$y(x) = \psi^{\ell}(x) \cdot \ell(x) + \psi^{k}(x) \cdot k(x), \tag{1}$$

where $\ell(x)$ is labor employed in task x, $k(x)$ is capital used in the production of task x, and $\psi^{\ell}(x)$ and $\psi^{k}(x)$ denote, respectively, the productivities of labor and capital in task x. We order tasks such that $\psi^{\ell}(x)/\psi^{k}(x)$ is nondecreasing and simplify the exposition by assuming that it is strictly increasing. We also suppose that when indifferent between producing a task with capital or labor, firms produce with capital. Therefore, there exists a threshold task θ such that tasks in [0, θ] are produced with capital and tasks in [θ, 1] are produced with labor. For now, there is no distinction between the adoption and the development of such technologies. We explore the implications of this distinction in section IV.B.

The household side is inhabited by a representative household that lives for two periods, $t = 0$ and $t = 1$. There is no production in period 0, but the representative household is endowed with $\bar{y}$ units of output. Out of this, it consumes c_0 and saves the remaining $k = \bar{y} - c_0$ units, which are allocated to producing capital. Capital is used during period 1, is subject to depreciation at the rate δ, and is rented to firms at the rental rate R, so that households earn an after-tax return of $(R - \delta) \cdot (1 - \tau^k)$. The period 1 budget constraint facing the household is

$$c \leq (1 + (R - \delta) \cdot (1 - \tau^k)) \cdot k + w \cdot (1 - \tau^{\ell}) \cdot \ell,$$

where R is the rental rate on capital paid by firms and w is the wage rate. Tax revenues are used for financing a fixed level of government expenditure, denoted by g.

The household chooses consumption and the supply of capital and hours to maximize

$$u(\bar{y}-k)+c-v(\ell).$$

Here, $u(\bar{y}-k)$ is a concave function representing the utility from consuming $\bar{y}-k$ units of output in period 0; c denotes the utility from consumption in period 1; and $v(\ell)$ is a convex function representing the disutility from working. Quasi linearity in period 1 is imposed for simplicity (see the online appendix for more general preferences).

We allow for various types of frictions in the labor market, modeled as introducing a wedge between the market wage and the representative household's marginal cost of supplying labor. We denote this wedge by $\varrho \geq 0$.[5]

Market clearing for capital and labor requires $k = \int_0^1 k(x)dx$ and $\ell = \int_0^1 \ell(x)dx$. To ensure uniqueness of optimal taxes below, we suppose that $u'(\bar{y}-k)\cdot k$ and $v'(\ell)\cdot\ell$ are convex. In addition, we assume that the equilibrium involves a positive net rate of return on investment. Finally, we denote by $\varepsilon^k(k)$ and $\varepsilon^\ell(\ell)$ the Hicksian elasticities of capital and labor. These are given by the response of capital and labor supply to a permanent percent change in the relevant keep rates (one minus the tax rates):

$$\varepsilon^k(k)=\frac{d\ln k}{d\ln(1-\tau^k)}=-\frac{u'(\bar{y}-k)-1}{u''(\bar{y}-k)\cdot k}\geq 0 \quad \text{and}$$

$$\varepsilon^\ell(\ell)=\frac{d\ln \ell}{d\ln(1-\tau^\ell)}=\frac{v'(\ell)}{v''(\ell)\cdot \ell}\geq 0.$$

As the equation for $\varepsilon^k(k)$ makes clear, the concavity of period –1 utility, $u(\bar{y}-k)$, ensures that the marginal rate of substitution between consumption today and tomorrow is increasing in k and thus the supply of capital is not perfectly elastic, otherwise $\varepsilon^k(k)$ would be infinite.[6]

5. As shown in Acemoglu, Manera, and Restrepo (in progress), this wedge can be derived from bargaining between workers and firms or from efficiency wage considerations.

6. A complementary reason for finite $\varepsilon^k(k)$ is that the technology for investment is convex (for example, the production of k units of capital requires $\phi(k)$ units of period 0 resources, where ϕ is strictly convex). If the profits from producing capital cannot be directly taxed, our optimal tax formulae apply regardless of whether $\varepsilon^k(k)$ reflects changes in the marginal rate of substitution between consumption today and tomorrow as a function of k or a convex investment technology.

Note that our formulation assumes that τ^k is a tax on net—after depreciation—returns, not gross returns, and our formula for $\varepsilon^k(k)$ computes it as the elasticity of capital to a percent change in one minus the net tax on capital.

I.B. Equilibrium

Given taxes $\{\tau^k, \tau^\ell\}$ and the labor wedge ϱ, a market equilibrium is defined by factor prices $\{w, R\}$, a tuple of current output, consumption, capital, and labor, $\{y, c, k, \ell\}$, and an allocation of tasks to factors, such that this allocation minimizes the after-tax cost of producing each task and the markets for capital, labor, and the final good clear. The online appendix shows that the equilibrium level of output can be represented as:

$$(2) \quad y = f(k, \ell; \theta) = \left(\left(\int_0^\theta \psi^k(x)^{\lambda-1} dx \right)^{\frac{1}{\lambda}} \cdot k^{\frac{\lambda-1}{\lambda}} + \left(\int_\theta^1 \psi^\ell(x)^{\lambda-1} dx \right)^{\frac{1}{\lambda}} \cdot \ell^{\frac{\lambda-1}{\lambda}} \right)^{\frac{\lambda}{\lambda-1}},$$

where the threshold task θ satisfies

$$(3) \quad \theta = \theta^m(k, \ell) \equiv arg\ max_{\theta \in [0,1]} f(k, \ell; \theta).$$

Moreover, factor prices are given by the usual marginal conditions $f_k = R$ and $f_\ell = w$. Consequently, the market-clearing condition for capital is

$$(4) \quad u'(\bar{y} - k) = 1 + (f_k - \delta) \cdot (1 - \tau^k),$$

while the market-clearing condition for labor is

$$(5) \quad v'(\ell) = f_\ell \cdot (1 - \varrho) \cdot (1 - \tau^\ell),$$

so that the wedge ϱ and the labor tax τ^ℓ distort the labor market in similar ways.

Finally, the government budget constraint takes the form

$$(6) \quad g \leq \tau^k \cdot (f_k - \delta) \cdot k + \tau^\ell \cdot f_\ell \cdot \ell.$$

A couple of points about this equilibrium are worth noting. As emphasized in Acemoglu and Restrepo (2018, 2019b), though the output level in the economy can be represented by a constant elasticity of substitution (CES) aggregate of capital and labor, the implications of this setup are very different from models that assume a CES production function with factor-augmenting technologies. First, there is a crucial distinction between capital intensity of production given a fixed allocation of tasks to factors and automation, represented by an increase in θ—which involves the substitution of capital for tasks previously performed by labor. This can be seen from the fact that holding the task allocation constant, the elasticity of substitution between capital and labor is λ, but when θ adjusts, the elasticity is greater. Second, further automation increases productivity but can easily reduce labor demand and the equilibrium wage because of the displacement it creates (mathematically, this works by changing the share parameters of the CES). In contrast, with a standard CES production function labor demand necessarily increases when capital becomes more productive. Third, and for the same reason, automation always reduces the labor share. Finally, our framework also clarifies that marginal increases in automation have second-order effects on aggregate output because, as shown in equation (3), the level of automation is chosen optimally.

I.C. Optimal Policy

We now characterize optimal policy by considering the choices of a benevolent social planner that sets capital and labor taxes τ^k and τ^ℓ and directly controls the extent of automation, represented by θ. We refer to the maximization problem of this planner as the Ramsey problem. As usual, this problem can be transformed so that the planner directly chooses an allocation $\{c, \ell, k, \theta\}$ that maximizes household utility subject to the resource constraint of the economy and a single implementability condition, which combines the government budget constraint in equation (6) and input market equilibrium conditions, equations (4) and (5):

$$(7) \quad \max_{c,\ell,k,\theta} \quad u(\bar{y} - k) + c - v(\ell)$$

$$\text{subject to: } c + g = f(k, \ell; \theta) + (1 - \delta) \cdot k \quad \text{(resource constraint)}$$

$$g \leq f(k, \ell; \theta) + (1 - \delta) \cdot k - u'(\bar{y} - k) \cdot k \frac{v'(\ell) \cdot \ell}{1 - \varrho}$$

$$\text{(implementability condition)}$$

Because the planner is assumed to choose the level of automation θ, we do not impose $\theta = \theta^m(k, \ell)$ as an additional constraint. We discuss issues of how the planner's choice of automation can be implemented below. Throughout, we use $\mu > 0$ to denote the multiplier on the implementability condition, which also gives the social value of public funds.

PROPOSITION 1: Optimal capital and labor taxes and automation

The unique solution to the Ramsey problem in equation (7) satisfies

$$(8) \qquad \frac{\tau^{k,r}}{1-\tau^{k,r}} = \frac{\mu}{1+\mu}\frac{1}{\varepsilon^k(k)} \quad \textit{and} \quad \frac{\tau^{\ell,r}}{1-\tau^{\ell,r}} = \frac{\mu}{1+\mu}\frac{1}{\varepsilon^\ell(\ell)} - \frac{\varrho}{1+\mu}$$

and $\theta^r = \theta^m(k, \ell)$.

The proof of this proposition, like those of all other results in this paper, is provided in the online appendix. The optimal tax formulas in equation (8) follow from the first-order conditions for the maximization problem in equation (7). Uniqueness follows from the fact that the Ramsey problem is convex (the objective function is quasi-concave and the constraint set is convex).

This proposition provides simple and intuitive formulas for the optimal taxes on capital and labor related to the social value of public funds and the inverse of the elasticity of supply of these factors. The formulas show that taxes should be lower for more elastic factors, and in addition, the optimal labor tax is further lowered by the presence of labor market frictions. This latter feature is intuitive: labor market frictions reduce employment beyond the socially optimal level, and the planner corrects for this by reducing labor taxation.

An immediate corollary of this proposition provides one set of sufficient conditions for uniform (symmetric) taxation of capital and labor—$\varepsilon^k(k) \simeq \varepsilon^\ell(\ell)$ and $\varrho \simeq 0$.

Corollary 1

If $\varepsilon^k(k) = \varepsilon^\ell(\ell)$) and $\varrho = 0$, uniform taxation of capital and labor is optimal.

In section III we will see that realistic values of these parameters are not too far from $\varepsilon^k(k) \simeq \varepsilon^\ell(\ell) > 0$, but labor market imperfections imply $\varrho > 0$, so that our framework yields lower labor taxes than capital taxes in the optimum.

Although the formulas in equation (8) apply in a two-period model, the online appendix shows that, under the key assumption that the government must run a balanced budget, these formulas extend to an infinite-horizon setting.[7] The online appendix also derives similar formulas for general preferences over consumption and leisure and clarifies the relationship

7. Even if the government is allowed to incur debt or accumulate assets, the result that the optimal tax system should not simultaneously impose significant taxes on labor and zero (or small) taxes on capital extends to an infinite-horizon setting provided that the long-run elasticity of capital supply, $\varepsilon^k(k)$, is not infinite. Straub and Werning (2020) show that in a

between our result and Atkinson and Stiglitz's (1972) principles of optimal commodity taxation.

In line with Diamond and Mirrlees (1971), proposition 1 also shows that, once optimal taxes are imposed on capital and labor, the planner has no reason to deviate from equilibrium automation decisions, $\theta^r = \theta^m(k, \ell)$. This is because any distortions in the labor market are corrected by optimal taxes, and thus, factor prices accurately reflect the social values of capital and labor. Consequently, profit-maximizing automation decisions are optimal as well. We will see that this is no longer true when taxes are not optimal or are subject to additional constraints.

I.D. Excessive Automation with Tax Distortions

Naturally, taxes in practice need not coincide with those characterized in proposition 1 both because of additional constraints and for political economy reasons (policymakers have other objectives and face political or other unmodeled economic constraints). When that is the case, either capital or labor taxes can be (relatively) too low. The interesting case for us, both for conceptual and empirical reasons, is the one where capital taxes are too low and labor taxes are too high, and the necessary and sufficient condition for this follows from equation (8) in proposition 1 and is presented in the next corollary.

Corollary 2

If the tax system $\{\tau^k, \tau^\ell\}$ is below the peak of the Laffer curve and satisfies

$$\frac{\dfrac{\tau^\ell}{1-\tau^\ell}+\varrho}{\dfrac{1}{\varepsilon^\ell(\ell)}-\dfrac{\tau^\ell}{1-\tau^\ell}} > \frac{\dfrac{\tau^k}{1-\tau^k}}{\dfrac{1}{\varepsilon^k(k)}-\dfrac{\tau^k}{1-\tau^k}}, \tag{9}$$

then $\tau^\ell > \tau^{\ell,r}$ and $\tau^{k,r} > \tau^k$—that is, the labor tax is too high and the capital tax too low.

The inequality expressed in equation (9) is sufficient for the tax system being biased against labor and in favor of capital.[8] An important

representative household economy where preferences are not time-additive separable and the tax system is not constrained by other considerations, optimal taxes on both capital and labor should converge to zero. We prove in the online appendix that if in addition there are labor market distortions, then optimal long-run taxes are lower on labor than capital.

8. The government budget constraint implies that both taxes cannot be too high or too low at the same time (provided that we are below the peak of the Laffer curve, meaning that tax revenues cannot be increased by lowering both taxes). Thus, equation (8) is sufficient for $\tau^\ell > \tau^{\ell,r}$ and $\tau^{k,r} > \tau^k$.

implication of such a biased tax structure is that there is too little employment relative to the optimal allocation in proposition 1, and thus marginal increases in employment will have first-order positive effects on welfare. We exploit this insight in the next proposition, where we take the tax system as given and consider a marginal change in automation. To do this in the simplest way, we relax the government budget constraint, equation (6), and value changes in revenue at the social value of public funds given by the multiplier μ.

PROPOSITION 2: When reducing automation improves welfare

Suppose that the tax system $\{\tau^k, \tau^\ell\}$ satisfies equation (9) (and is thus biased against labor and in favor of capital). Welfare (inclusive of fiscal costs and benefits) increases following a small reduction in θ below $\theta^m(k, \ell)$. A small reduction in θ also increases net output provided that $\varepsilon^\ell(\ell) > \varepsilon^k(k)$ and government revenue provided that $\tau^\ell \cdot (1 + \varepsilon^\ell(\ell)) > \tau^k \cdot (1 + \varepsilon^k(k))$.

This result shows that, in contrast with proposition 1, when taxes are not optimal and are biased against labor (in the sense that equation (9) holds), it is welfare improving to restrict automation below its equilibrium level. This result is intuitive in light of the observation in corollary 2 that employment is below the socially optimal level. Specifically, a small reduction in automation will create a first-order welfare gain by shifting demand from capital to labor. Distorting automation is costly, but starting from the equilibrium level of automation, $\theta^m(k, \ell)$, this cost is second-order, since $f_\theta(k, \ell; \theta^m(k, \ell)) = 0$, and hence, a small reduction in automation is welfare improving. This intuition relates proposition 2 to the notion of "so-so (automation) technologies" proposed in Acemoglu and Restrepo (2019a, 2019b): automation is not beneficial to labor when it only increases productivity by a small amount, while still creating the usual displacement of workers as tasks are reallocated from them to capital. The equilibrium condition $f_\theta(k, \ell; \theta^m(k, \ell)) = 0$ implies that automation technologies adopted at marginal tasks are, by definition, so-so. The planner is therefore happy to sacrifice these so-so technologies in order to help labor.[9]

As we will see in section III, the US tax system lies within the range that satisfies equation (9), so that there are prima facie reasons for suspecting that the level of automation may be excessively high in the US economy, as in this proposition.

9. If automation decisions were constrained by available technology, that is, θ had to be less than some $\bar{\theta} < 1$ as in Acemoglu and Restrepo (2018, 2019a), we could have that $f_\theta(k, \ell; \theta^m(k, \ell)) > 0$ if $\theta^m(k, \ell) = \bar{\theta}$. In this case, productivity gains from automating marginal tasks could be positive. If they were sufficiently large, then automation would no longer be a so-so technology and proposition 2 would not apply.

One common intuition is that when confronted with a tax system with distortions, $\{\tau^k, \tau^\ell\}$, the best policy is to redress these tax distortions directly. We next show that this is not always the case. In particular, if for other reasons taxes on labor cannot be reduced below a certain threshold (which we denote by $\overline{\tau}^\ell$), then the tax system satisfies equation (9) and is biased against labor, but this does not necessarily imply that capital taxes should be increased. Rather, constrained optimal policy calls for a reduction in the equilibrium level of automation and may even involve a *lower* tax on capital. Before presenting this result, let us note that in this case we are imposing $\tau^\ell \geq \overline{\tau}^\ell$, which can be expressed as an additional constraint on the Ramsey problem in equation (7) of the form

$$v'(\ell) \leq (1 - \overline{\tau}^\ell) \cdot (1 - \varrho) \cdot f_\ell, \tag{10}$$

where the lower bound on labor taxes translates into an upper bound on the marginal disutility from work. In the next proposition, we denote the multiplier on this constraint by $\gamma^\ell \cdot \ell \geq 0$ (where the ℓ simply normalizes the multiplier and makes it easier to interpret).

PROPOSITION 3: Excessive automation with tax distortions

Consider the constrained Ramsey problem of maximizing equation (7) subject to the additional constraint $\tau^\ell \geq \overline{\tau}^\ell$, and suppose that in the solution to this problem equation (10) binds. Then the constrained optimal taxes and allocations are given by a labor tax of $\tau^{\ell,c} = \overline{\tau}^\ell$ and a tax subsidy on capital that satisfies

$$\frac{\tau^{k,c}}{1 - \tau^{k,c}} = \frac{\mu}{1 + \mu} \frac{1}{\varepsilon^k(k)} - \frac{\gamma^\ell}{1 + \mu} \cdot (1 - \overline{\tau}^\ell) \cdot (1 - \varrho) \frac{f_{\ell k} \cdot \ell}{u' - 1} \tag{11}$$

and a level of automation $\theta^c < \theta^m (k, \ell)$.

Before discussing the implications of this proposition, we explain the meaning of the constraint expressed in equation (10). The fact that this constraint is binding means that the planner would have chosen a tax rate on labor $\tau^{\ell,r}$ below $\overline{\tau}^\ell$, but the constraint forces the planner to set a higher tax on labor of $\overline{\tau}^\ell$, which results in a tax system biased against labor and in favor of capital, or in other words, the inequality in equation (9) holds. This also implies that the level of employment is below what the planner would have chosen in the unconstrained Ramsey problem.

Given this biased tax system, the planner wants automation to be less than its equilibrium level. The intuition is identical to that in proposition 2: the reduction in automation creates a second-order productivity cost but a first-order gain via its impact on increased employment. Importantly, this holds even when capital taxes can be freely adjusted.

Moreover, the optimal capital tax formula in equation (11) has an additional negative term on the right-hand side relative to equation (8). This negative term can lead not just to lower capital taxes than in the unconstrained Ramsey problem in proposition 1, but even to capital subsidies.[10] The combination of lower capital taxes and limiting the set of tasks that are automated ensures that capital gets used intensively in tasks that are (and should be) automated, while avoiding the displacement of workers from marginal tasks. Both of these margins contribute to raising employment, which increases welfare when the tax system is biased against labor. This is related to the discussion of deepening of automation in Acemoglu and Restrepo (2019a): deepening of automation, which means an increase in the use or productivity of capital in tasks that are already automated, is always beneficial for labor. What is potentially damaging to labor is the extensive margin of automation—because this displaces workers from tasks they were previously performing. Proposition 3 builds on this logic: the planner would like to reduce the range of tasks that are automated by reallocating marginal tasks back to labor and may also want to reduce capital taxes or even subsidize capital at the same time, so that automated tasks can use capital more intensively.

Proposition 3 focused on the case with a lower bound on labor taxes. An equally plausible case is one where, because of political influence of capital owners or because of concerns about capital flight, there is an upper bound on capital taxes.[11] Proposition 2 in the online appendix establishes that in this case, too, the planner would like to reduce automation below its market level, even if taxes on labor can be adjusted. The intuition is similar: the upper bound on capital taxation leads to a tax system biased in favor of capital and against labor, and this makes the displacement of labor by capital in marginal tasks socially costly.

I.E. Implementation

To ease exposition, we have so far assumed that the planner can directly control θ. We now discuss how the desired level of θ can be implemented via taxes. Recall that $k(x)$ is the capital used in task x, and so far we have

10. This might at first appear surprising, especially because the program in proposition 1 is convex, so moving in the direction of the unconstrained optimum should be beneficial. However, the convexity is in the space of allocations and does not imply convexity in the space of taxes. Therefore, increasing the tax rate on capital toward $\tau^{k,r}$ is not necessarily welfare-improving.

11. A similar constraint on capital taxation is used in the optimal taxation literature (Chamley 1986; Judd 1999; Straub and Werning 2020).

assumed that all types of capital are taxed at the same uniform rate, τ^k. In practice, taxes vary by type of capital (e.g., equipment, software, structures) and industry (because of differential depreciation allowances). In the context of our model, this can be viewed as a task-specific capital tax rate of $\tau^k(x)$. The next proposition establishes when such task-specific capital tax rates are useful and in the process further clarifies the nature of optimal policy interventions.

PROPOSITION 4: Automation tax

Suppose the planner can set task-specific capital taxes and cannot directly control automation decisions. Then, first, under the conditions of proposition 1, the planner sets a uniform capital tax rate, that is, $\tau^k(x) = \tau^k$. Second, under the conditions of proposition 3, the planner prefers to depart from uniform capital taxation. In particular, the planner can implement the level of automation $\theta^c < \theta^m(k, \ell)$ with the following tax scheme:

$$\tau^k(x) = \begin{cases} \tau^k & \text{for } x \leq \theta^c \\ \tau^k + \tau^A & \text{for } x > \theta^c \end{cases},$$

where $\tau^A > 0$ is a task-specific automation tax.

The reason (unconstrained) optimal policy has no use for task-specific taxes is intuitive: in the unconstrained Ramsey problem, there is no need to distort equilibrium automation decisions. However, in the presence of additional constraints, the planner would like to reduce automation to $\theta^c < \theta^m(k, \ell)$, and the planner can achieve this by imposing an incremental tax to capital used in tasks above θ^c. By design, these incremental taxes encourage the use of capital in tasks where capital has a comparative advantage (which helps labor via complementarities across tasks) and discourages the automation of marginal tasks (which also benefits labor by preventing its displacement). In what follows, we refer to the incremental tax on capital τ^A as an automation tax.

II. The US Tax System

In this section, we first introduce the notion of *effective taxes* on capital and labor. Effective taxes summarize the average distortion that the US tax system introduces in the use of capital and labor. We then provide formulas for effective taxes that take into account the various elements of the US tax code and their interaction with the type of financing and ownership structure of the firm making investment decisions.

II.A. Defining Effective Taxes on Capital

In our framework, τ^k is the effective tax on (the use of) capital. It is defined as the wedge that the tax system introduces between the internal rate of return for a firm investing in capital and the after-tax rate of return paid to investors. The US tax system includes several taxes, not just a single effective tax on the use of capital. We have personal income taxes on capital income, corporate income taxes, depreciation allowances, and many other instruments that contribute to taxes on different types of capital. Moreover, these taxes vary by form of organization (C corporation versus pass-through) and type of financing (equity versus debt).[12]

We start by providing formulas for effective taxes on the use of capital by type of asset, j, form of organization, and type of financing. To simplify the exposition, we assume the economy is in steady state—the capital-labor ratio remains constant, the tax system is not expected to change, the price of capital goods changes at a constant rate $\pi^j = q_t^j/q_{t-1}^j$, and the capital stock of type j depreciates at a constant rate $\delta^j > 0$.

The internal rate of return of investing one dollar in equipment j at time $t - 1$ is given by

$$r^{f,j} = \text{mpk}^j - \overline{\delta}^j,$$

where mpk^j is the marginal product of investing one dollar in asset j and $\tilde{\delta}^j = 1 - \pi^j \cdot (1 - \delta)$ denotes the total depreciation of the asset (inclusive of investment price changes). Let us denote the after-tax steady-state rate of return to investors by r. The effective tax rate on capital of type j, $\tau^{k,j}$, can then be defined as

$$\frac{1}{1-\tau^{k,j}} = \frac{r^{f,j}}{r} = \frac{\text{mpk}^j - \tilde{\delta}^j}{r}. \tag{12}$$

This formula aligns closely with the effective capital taxes in our conceptual framework in the previous section. In particular, in equation (4), $\frac{1}{1-\tau^k}$ is equal to the wedge (ratio) between the return to the firm from

12. Pass-through organizations include both S corporations and other pass-throughs, such as sole proprietor businesses and partnerships, and are subject to different tax rules, as we explain below.

using capital—$\text{mpk}^j - \tilde{\delta}^j$ here and given by $f_k - \delta$ in equation (4)—and the return demanded by investors— r here and $u'(\bar{y} - k) - 1$ in equation (4).[13]

The computation of effective tax rates requires measuring the marginal product of capital. We follow Hall and Jorgenson (1967) and back out the marginal product of capital using a representative firm's first-order condition for investment. We need to distinguish between C corporations and pass-through businesses as well as the source of financing, since each of these combinations implies a different first-order condition for investment as well as a different set of taxes on the income generated from capital.

For C corporations that finance their investment with equity, the first-order condition is

$$\text{mpk}^j = \frac{1 - \alpha^j \cdot \tau^c}{1 - \tau^c} \cdot \left(r^e + \tilde{\delta}^j\right), \tag{13}$$

where τ^c is the corporate income tax rate and $\alpha^j \in 0, 1$ are discounts from depreciation allowances, which reduce taxable income and are discussed in the next subsection. In the absence of corporate income taxes, this expression is identical to the standard user cost formula. In addition, r^e is the pretax return to equity holders. This implies that $r = r^e \cdot (1 - \tau^{e,c})$, where $\tau^{e,c}$ is the tax rate on income resulting from ownership of public equity.

Combining the formula for effective taxes in equation (12) with the first-order condition for investment in equation (13), the effective tax rate for an equity-financed C corporation is

$$\frac{1}{1 - \tau^{k,j}_{\text{c-corp,equity}}} = \frac{1}{1 - \tau^{e,c}} \cdot \left(\frac{r^e + \tilde{\delta}^j}{r^e} \frac{1 - \alpha^j \cdot \tau^c}{1 - \tau^c} - \frac{\tilde{\delta}^j}{r^e} \right). \tag{14}$$

The formula shows that the effective tax on capital depends on the taxation of capital income of equity owners, corporate income tax rates, and depreciation allowances. It reiterates that depreciation allowances can significantly offset corporate taxes. For example, with full (immediate) expensing, which corresponds to $\alpha^j = 1$, we would have $\tau^{k,j}_{\text{c-corp,equity}} = \tau^{e,c}$.

The main difference for pass-through businesses is that these organizations do not pay the corporate income tax and are only subject to

13. An alternative is to use a formula for effective taxes based on gross returns: $\frac{1}{1 - \tau^{k,j}_{\text{gross}}} = \frac{\text{mpk}^j}{r + \tilde{\delta}^j}$. All of our results can be expressed in terms of gross returns, but this would require adjusting the empirical estimates of capital supply elasticities, which are in terms of net returns.

personal income taxation. Depreciation allowances in this case lower personal income tax obligations for business owners. The formula for the effective tax on the use of capital for a pass-through business that is financing its investment with (private) equity is

$$\frac{1}{1-\tau^{k,j}_{\text{passthrough,equity}}} = \left(\frac{r^e + \tilde{\delta}^j}{r^e} \frac{1-\alpha^j \cdot \tau^{o,p}}{1-\tau^{o,p}} - \frac{\tilde{\delta}^j}{r^e} \right), \tag{15}$$

where $\tau^{o,p}$ denotes the individual tax rate on the income of owners of pass-through businesses. Note again that with immediate expensing ($\alpha^j = 1$), we have $\tau^{k,j}_{\text{passthrough,equity}} = 0$.

We next turn to debt-financed investments, which allow a further tax discount by subtracting interest payments from taxable income. The presence of these additional tax discounts modifies the first-order condition for investment to

$$\text{mpk}^j = \frac{1-\alpha^j \cdot \tau^c}{1-\tau^c} \cdot \left(r^b \cdot (1-\tau^c) + \tilde{\delta}^j \right), \tag{16}$$

where r^b is the return offered to bondholders and the term $r^b \cdot (1 - \tau^c)$ captures the fact that tax liabilities are lower because of the deduction of interest rate payments. Note that the after-tax return to households that own bonds is $r = r^b \cdot (1 - \tau^{b,c})$, where $\tau^{b,c}$ is the personal income tax rate for capital income from C corporation bonds.

Combining the formula for effective taxes in equation (12) with the first-order condition for investment in equation (16), the effective tax rate for a debt-financed C corporation is

$$\frac{1}{1-\tau^{k,j}_{\text{c-corp,debt}}} = \frac{1}{1-\tau^{b,c}} \cdot \left(\frac{r^b \cdot (1-\tau^c) + \tilde{\delta}^j}{r^b} \cdot \frac{1-\alpha^j \cdot \tau^c}{1-\tau^c} - \frac{\tilde{\delta}^j}{r^b} \right). \tag{17}$$

The effective tax on capital again depends on the personal income tax rate of bondholders, corporate income tax rates, interest rate deductions, and depreciation allowances.[14] The additional tax discounts can easily lead

14. Our model assumes that new and underappreciated old (already installed) capital are perfect substitutes and thus face the same tax rate. When new and old capital are imperfect substitutes, bonus depreciation allowances and other deductions will make new capital cheaper relative to already installed capital (see Auerbach and Kotlikoff 1987). To the extent that capital involved in automation tends to be new capital, this would create an additional incentive for excessive automation.

to a net subsidy to the use of capital. In particular, with full expensing ($\alpha^j = 1$), we have $\tau^{k,j}_{\text{c-corp,debt}} \approx \tau^{b,c} - \tau^c$, which is negative if bondholders face lower individual tax rates than corporations.

Owners of pass-through businesses can also subtract their interest payments on debt from their taxable income. However, if they issue bonds, payments to bondholders are subject to personal income taxation. The formula for the effective tax on the use of capital for a pass-through business that is financing its investment with debt is similar to that of a C corporation and given by

$$\text{(18)} \qquad \frac{1}{1-\tau^{k,j}_{\text{passthrough,debt}}} = \frac{1}{1-\tau^{b,p}} \cdot \left(\frac{r^b \cdot (1-\tau^{o,p}) + \tilde{\delta}^j}{r^b} \frac{1-\alpha^j \cdot \tau^{o,p}}{1-\tau^{o,p}} - \frac{\tilde{\delta}^j}{r^b} \right),$$

where $\tau^{b,p}$ denotes the individual income tax rate applying to holders of pass-throughs' bonds. As before, with full expensing ($\alpha^j = 1$), we would have $\tau^{k,j}_{\text{passthrough,debt}} \approx \tau^{b,p} - \tau^{o,p}$, which is negative if bondholders face lower income taxes than owners of pass-through businesses.

II.B. Computing Effective Taxes on Capital

We compute effective taxes for equipment, software, and structures separately. For each type of capital good, we calculate effective taxes by form of organization and type of financing, and we aggregate these taxes into a single effective tax rate for the relevant type of capital using investment shares as weights. The online appendix details the sources and numbers used in our calculations. Here we outline the computations of the key objects in our formulas for effective taxes on capital: depreciation allowances, α^j; corporate income taxes and taxes on owners of equity and pass-throughs; and interest rates, economic depreciation, and investment prices.

DEPRECIATION ALLOWANCES The tax discount term, α^j, is equal to the present discounted value of depreciation allowances associated with one unit of capital purchased at time t, which can be computed as

$$\text{(19)} \qquad \alpha^j = d^j_0 + \sum_{s=0}^{\infty} d^j_{s+1} \cdot \prod_{\tau=0}^{s} \frac{1-d^j_\tau}{1+r},$$

where d^j_s denotes the fraction of the investment that a firm is allowed to subtract from its tax liabilities s years after the purchase.

One useful benchmark is when firms can subtract the economic depreciation of their capital goods each period. In the above formula, this means

$d_0^j = 0$ and a constant depreciation rate of δ^j from there on, which adds up to an allowance of $\tilde{\alpha}^j = \delta^j/(\delta^j + r) < 1$.

The Internal Revenue Service (IRS) and the US tax code handle depreciation allowances quite differently from this benchmark, however. The way in which depreciation allowances are determined is specified in IRS Publication 946. The current system places each type of capital under a specific class life—the number of years that a new unit of capital lasts for tax purposes—based on its characteristics and sector. The first reason why tax discounts α^j differ from the one given by constant economic depreciation, $\tilde{\alpha}^j$, is that the depreciation rate implied by a class life is different from the economic depreciation rate.

A second source of an additional tax discount is that the tax code requires taxpayers to follow specific depreciation schedules and enables front-loading of allowances. When computing their tax discount, firms may use a combination of straight line and declining balance methods that yields the highest possible discount. The straight line method allows firms to expense a constant fraction of their initial investment (or undepreciated investment in the initial year in which the method is applied) for each year of remaining tax life. The declining balance method can be used for assets with a class life below twenty years and allows firms to front-load their depreciation allowances by expensing a decreasing fraction of their initial investment each year. Assets in a class life of ten years or less can be depreciated using a 200 percent declining balance rule, which allows firms to expense their undepreciated investment at two times the rate prescribed by the straight line method (2×10 percent for an asset in a class life of ten years). Firms can then switch to the straight line method near the end of the asset life to maximize their allowances.[15] Assets with a class life

15. As an example, consider the allowances generated by the purchase of a machine with a class life of ten years. Suppose the purchase takes place in the middle of the year. The straight line method allows a deduction of 5 percent of the cost in the first year, 10 percent for the following nine years, and 5 percent on the eleventh year. The 200 percent declining balance method gives an allowance of 10 percent in the first year (two times the straight line rate of 5 percent), 18 percent in the second year (two times the straight line rate of 10 percent times the undepreciated stock, 90 percent), 14.4 percent in the third year (two times the straight line rate of 10 percent times the undepreciated stock, 72 percent). This continues up to year seven, where the method prescribes an allowance of 5.89 percent, which is below the straight line method allowance of 6.55 percent computed on the undepreciated stock of capital and four and a half years of useful life left. Therefore, the schedule for ten-year property follows the 200 percent declining balance method until year seven and switches to a constant allowance of 6.55 percent of the undepreciated cost for the remaining four and a half years. For further discussion and examples, see the appendix in House and Shapiro (2008).

between ten and twenty years, on the other hand, can be depreciated using a 150 percent declining balance rule, while assets with a class life of more than twenty years adhere to the straight line method.

The third and final source of large discounts from depreciation allowances is recent changes in legislation, passed as part of economic stimulus plans, which introduced bonus depreciation.[16] Under current bonus depreciation provisions, most capital with a class life below twenty years enjoys a 100 percent bonus depreciation, meaning that investors can immediately expense their capital purchases as current costs. This full expensing yields the maximum discount of $\alpha^j = 1$.[17]

We compute α^j_t for 1980–2018 for each type of capital, taking into account changes in the treatment of depreciation allowances and bonus depreciation programs (excluding the further reductions in effective capital taxes generated by the 2017 tax reforms, since these did not affect the automation decisions throughout the 2010s). When computing α^j_t, we assume that firms anticipate no future changes in the tax code, so that they expect current rates to apply in the future.[18]

Figure A.4 in the online appendix plots $\tilde{\alpha}^j$ and α^j for software, equipment, and nonresidential structures. The figure shows that α^j typically exceeds $\tilde{\alpha}^j$ for software and equipment and that recent bonus depreciation provisions generated an increase in allowances, bringing α^j close to 1 for software and equipment in the 2010s.

TAX RATES ON CORPORATIONS AND CAPITAL OWNERS Effective taxes on capital also depend on taxes on corporations and the households who own

16. In particular, the Job Creation and Worker Assistance Act of 2002 (JCWAA) introduced a 30 percent bonus depreciation for 2002–2003; the Jobs and Growth Tax Relief Reconciliation Act of 2003 (JGTRRA) raised the bonus to 50 percent for 2004; the Economic Stimulus Act of 2008 introduced a 50 percent bonus, extended until 2017 by successive bills; the Tax Relief, Unemployment Insurance Reauthorization, and Job Creation Act of 2010 temporarily raised the bonus to 100 percent (full expensing) between September 2010 and the end of 2011. Finally, the Tax Cuts and Jobs Act of 2017 raised bonus depreciation to 100 percent for 2018–2022.

17. A 100 percent bonus depreciation corresponds to $d_0 = 1$ and $d_s = 0$ for all s > 0 in equation (19). As stated above, capital allowances are generally set by the schedules in IRS Publication 946, which give a specific d^j_s for all $s, j,$ such that $\sum_{s=0}^{T_j} d^j_s = 1$, for each investment type j, and where T_j is the class life for the capital type j. When bonus depreciation is $\gamma < 1$, the taxpayer obtains a first-year bonus allowance equal to γ and then follows the schedules for depreciation allowances for the undepreciated capital stock. Therefore, the bonus allowance series, d^j_s, has $\tilde{d}^j_s = (1-\gamma)d^j_s$, for all $s \geq 1$, and $\tilde{d}^j_s = \gamma + (1-\gamma)d^j_0$ in the initial period.

18. Anticipated tax reforms create a reevaluation effect for capital that is already installed.

capital. We approximate the average marginal corporate income tax rate τ_t^c for each year as the average tax paid by C corporations:

$$\tau_t^c = \frac{\text{corporate tax revenue}}{\text{net surplus of C corporations}}.$$

The corporate tax revenues are obtained from National Income and Product Accounts (NIPA) tables. The computation of the tax base is presented in the online appendix. We start with operating surplus from corporations and subtract depreciation allowances. We then allocate a fraction of these profits to C corporations using data from the IRS on profits by type of corporation. The remaining share is accounted for by S corporations, which do not pay corporate income taxes, and this share is not included as part of the tax base in the above calculation. The share of corporate profits generated by C corporations has fallen over time from 93 percent in 1980 to 61 percent in 2018, in line with the findings of Smith and others (2019). Our calculations show that once we account for this changing share, the average tax rate on C corporations increased from 25 percent in 1981 to 35 percent in 2000 and then declined to 17.5 percent in 2018.

Note that we are computing corporate income taxes as an average of the taxes paid, rather than by using the statutory rate (46 percent in 1981, 35 percent in the intervening years, and 21 percent in 2018).[19] This is because many corporations pay less corporate income tax than implied by the statutory rate. Throughout, we interpret average taxes as averages of marginal tax rates faced by different types of firms.

Besides taxes paid by corporations, taxes paid by households on their capital income from equity and lending also contribute to the effective tax on the use of capital—the terms $\tau^{e,c}$, $\tau^{b,c}$, and $\tau^{b,p}$ in equations (14), (17), and (18). We compute $\tau^{e,c}$ as the average tax rate paid by owners of equity on their dividends and capital gains. We start by computing the share of corporate equity that is directly held by US households and is thus subject to taxation. Using data from the Board of Governors of the Federal Reserve System, we approximate this as the share of corporate equity owned by US households and nonprofit organizations serving these households, which has fallen from 58 percent in 1981 to 37 percent in 2018. We follow the

19. See the IRS SOI Tax Stats—Historical Table 24, https://www.irs.gov/statistics/soi-tax-stats-historical-table-24.

CBO (2014) and assume that the remaining share is owned by funds or kept in accounts that are not subject to additional taxation.

Taxes paid by households depend on how corporate profits are realized. Qualified dividends or capital gains are taxed at a maximum capital gains tax rate specified by the IRS.[20] These include dividends on stocks held for more than sixty-one days or capital gains on stocks owned for over a year. Ordinary (nonqualified) dividends or capital gains apply to stock owned over shorter periods and are taxed at the same rate as individual income. The remaining profits are for stocks held until death, whose capital gains are never realized and thus face no taxation. We compute the share of profits realized through ordinary dividends and short-term capital gains by using data from the IRS Individual Complete Report (Publication 1304, table A) for the period 1990–2017 and the IRS Statistics of Income (SOI) Tax Stats (Sales of Capital Assets Reported on Individual Tax Returns) for the period 1990–2012. Publication 1304 reports households' ordinary dividend income from corporate stocks, while the SOI Tax Stats reports the short-term capital gains on corporate stocks. Short-term dividends and ordinary capital gains account for the bulk of realized profits from C corporations (about 60 percent in recent years). The remaining share of profits is accounted for by long-term qualified gains and dividends, or by stocks held until death whose capital gains are never realized. We assume that each of these two forms makes up an equal share of profits, which aligns with what the CBO reports for 2011.

The average tax rate on profits derived from C corporation profits (after paying corporate taxes) is thus given by

$$\tau_t^{e,c} = \begin{array}{c}\text{share directly}\\ \text{owned}_t\end{array} \cdot \left(\begin{array}{c}\text{share short-}\\ \text{term ordinary}_t\end{array} \cdot \tau_t^{o} + \begin{array}{c}\text{share long-}\\ \text{term qualified}_t\end{array} \cdot \tau_t^{q} + \begin{array}{c}\text{share held}\\ \text{until death}_t\end{array} \cdot 0\% \right).$$

Here, τ_t^o is the average tax rate on short-term ordinary capital gains and dividends, and τ_t^q is the average tax rate on long-term qualified capital

20. The maximum capital gains tax rate is specified in IRS Publication 550. In 2018, taxpayers facing a marginal tax rate below 15 percent had a maximum capital gains rate of 0 percent. Taxpayers facing a marginal tax rate between 22 percent and 35 percent had a maximum capital gains tax rate of 15 percent. Finally, taxpayers facing a marginal tax rate of 35 percent faced a maximum capital gains tax rate of 20 percent.

gains and dividends. Both average taxes are computed using data from the Office for Tax Analysis for 1980–2014. In recent years, the average tax rate on ordinary short-term gains and dividends was $\tau_t^o = 24$ percent and the average tax rate on long-term qualified capital gains and dividends was $\tau_t^q = 18$ percent. Our estimates show that $\tau_t^{e,c}$ has hovered around a historical average of 15 percent and experienced a temporary reduction to 12.5 percent during the 2000s.[21]

Turning to taxation of rental income for bondholders, the CBO estimates that 52.3 percent of C corporation bonds are held directly by households, 14.9 percent generate income that is temporarily deferred for tax purposes, and the rest is held by funds or kept in accounts that are not subject to additional taxation. For pass-through entities, the share owned by households is larger, 76.3 percent, and the share deferred is 10.1 percent. Moreover, the CBO reports that the rental income owned by households is subject to personal income taxes at the average rate 27.4 percent in 2014. Supposing that temporarily deferred income is subsequently taxed at the same rate as the rest of rental income, we estimate the average tax paid by bondholders on their rental income from C corporations and pass-throughs, respectively, as $\tau^{b,c} = 16.84$ percent and $\tau^{b,p} = 23.25$ percent, and we assume that these rates have remained constant over time.

The final item required for our calculations is the tax rate paid by owners of pass-throughs, which we separate into S corporations and other pass-throughs (sole proprietor businesses and partnerships). Profits from S corporations are taxed at the individual income rate of the owners. We assume that the average tax rate paid by owners of S corporations is the same as the average tax paid by individuals earning ordinary short-term dividends and capital gains, τ_t^o.[22] In economic terms, this requires owners of S corporations to have a similar income profile as investors in public equity. In addition, part of the profits generated by S corporations accrue

21. Our estimate for $\tau^{e,c}$ assumes that new investments are financed with new equity. When new investments are partly financed with retained earnings, the effective tax on capital might be even lower. In addition, accrual-equivalent taxes on capital will be generally lower than the average tax rate we use for capital gains, because gains are typically postponed relative to accrual. We thank Alan Auerbach for raising these two issues.

22. Profits from S corporations are also taxed as corporate income by some states. To account for these taxes, we add the average state and local tax rate on businesses, which we compute by dividing state and local revenues from business taxes by the net operating surplus of corporations. State and local taxes on businesses are small in practice, with an average value near 3 percent in recent years.

only when the company is sold, and these profits are taxed at the maximum qualified rate, τ_t^q. Thus, we measure the average tax paid by owners of S corporations on their profits as

$$\tau^{o,s} = \tau_t^o - \text{share capital gains} \cdot (\tau_t^o - \tau_t^q).$$

Using data on sales of pass-through businesses reported by the IRS for 1990–2000, we estimate the average share of capital gains in S corporation profits as 25 percent and assume it has remained at this level over time. Our estimates imply that $\tau_t^{o,s}$ has been roughly constant as well, at about 27 percent, reaching 28 percent in 2018. Since self-proprietors' and partnerships' income is reported as personal income, we have no data on the tax rate faced by owners on profits, and so we assume that they face the average tax rate on income (obtained from the IRS, SOI Tax Stats), which has been approximately 14.6 percent in recent years.

Overall, our estimates imply that in 2011 the average corporate income tax was 26.4 percent (with equity holders paying an additional 11.8 percent on top of this), the average tax rate paid by S corporation owners was 23 percent, and the average tax rate paid by owners of other pass-throughs was 14.6 percent. These numbers align closely with those from the CBO and with Cooper and others (2016).[23]

INTEREST RATES, DEPRECIATION, AND INVESTMENT PRICES We assume a constant interest rate, a constant growth rate for investment prices, and a constant rate of economic depreciation for each asset that match historical averages from 1981 to 2017. We use a constant value of $r^b = 4.21$ percent per annum for bondholders, given by the average of the Moody's Seasoned AAA Corporate Bond Yield minus realized inflation between 1981 and 2017. Likewise, we use a constant value of $r^e = 4.36$ percent per annum for equity holders, which is the historical average of the real rate of return on the S&P 500 over 1957–2018. The constant growth rate for investment prices is estimated from the average change of investment price indexes by type of capital from the Bureau of Economic Analysis (BEA) fixed asset tables between 1981 and 2017. These imply an annual average growth rate of prices equal to −1.6 percent for software, −1 percent for equipment, and 2 percent for nonresidential structures. The economic depreciation rates, the δ_t^j's, are taken directly from the BEA fixed asset tables as the averages

23. Using IRS data, Cooper and others (2016) estimate that in 2011 C corporations paid an average tax rate of 23 percent (plus 8.25 percent on the household side), S corporations paid an average tax rate of 25 percent, and other pass-throughs paid an average tax rate of about 14.7 percent.

for 1981–2017 (the average economic depreciation rate per annum is 23.4 percent for software, 13.9 percent for equipment, and 2.6 percent for nonresidential structures).

II.C. Effective Taxes on Labor

In our model, τ^{ℓ} is the effective tax on (the use of) labor. However, as with capital, there is no single tax on labor in the US tax code. Instead, labor income is subject to a number of different taxes both at the federal and local levels. Means-tested public programs may generate additional implicit taxes on labor. The effective tax on labor is given by the wedge that the tax system introduces between the marginal product of labor and the before-tax wage, mpl^{f}. The representative firm will demand labor until the marginal product of labor, mpl^{f}, equals the cost of one unit of labor given by total compensation. That is,

$$\mathrm{mpl}^{f} = \text{compensation} = \text{salary} + \text{benefits}.$$

Wage income is subject to personal income tax at a rate τ^{h} and payroll taxes at a rate τ^{p}. Benefits are not taxed but might be imperfectly valued by workers, which we capture by converting them to an income-equivalent amount by multiplying them with $\varphi \in 0, 1$. Consequently, the after-tax return to work for the household is given by

$$w = \text{salary} \cdot (1 - \tau^{h} - \tau^{p}) + \text{benefits} \cdot \varphi.$$

The effective tax rate on labor is defined, analogously to the effective tax on capital, as

$$\frac{1}{1-\tau^{\ell}} = \frac{\mathrm{mpl}^{f}}{w} \Rightarrow \tau^{\ell} = \frac{\text{salary} \cdot (\tau^{h} + \tau^{p}) + \text{benefits} \cdot (1-\varphi)}{\text{compensation}}.$$

We measure the terms in this expression as follows. From national accounts we obtain data on salaries and total compensation for the corporate sector. We treat employers' contributions to pensions and health insurance as part of the benefits since these are not taxed. We assume that workers outside the corporate sector receive a similar split between benefits and salaries and are therefore subject to the same effective taxes. We use a payroll tax rate of 15.3 percent, which is the statutory rate that applies to all earners with an income below $132,900 in 2018 (a level that roughly matches the 95th percentile of income). Since the vast majority of jobs at risk of automation are performed by workers in the middle of the income distribution, the payroll tax of 15.3 percent is relevant for automation decisions and is incorporated into our effective tax rate on labor. We measure the personal income tax rate τ^h, consistent with our treatment of payroll taxes, as the average income tax paid by earners below the 95th percentile. This is computed from publicly available data from the IRS for 1986–2017. The estimate for τ^h has been stable in recent years at a level close to 10 percent.[24] Finally, we use a value of $\varphi = 0.65$ building on estimates from Gruber and Krueger (1991), Goldman, Sood, and Leibowitz (2005), and Lennon (2020), which suggest that one dollar of spending on benefits is valued on average at 65 cents by households. This increases our estimates for τ^ℓ by 3 percent.

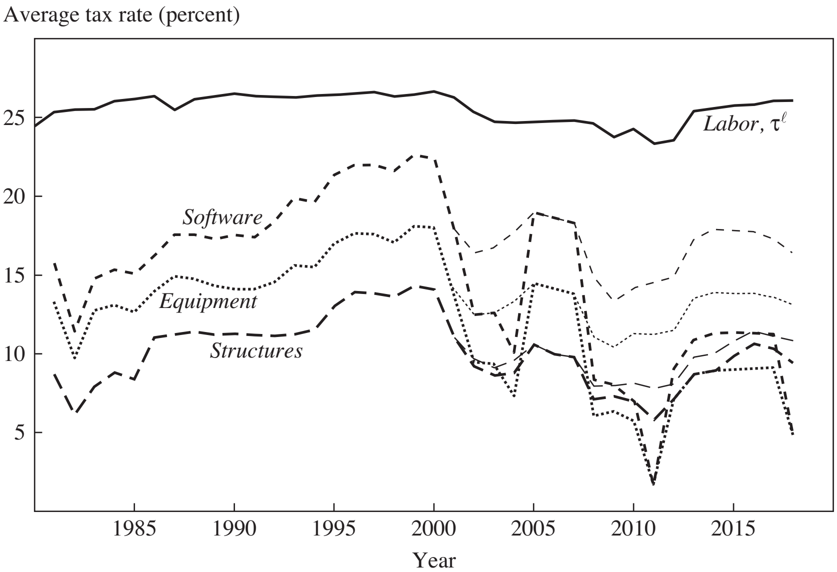
Besides our baseline estimate for τ^ℓ described above, in the online appendix we present results using an estimate for the effective tax on labor which incorporates the implications of means-tested welfare programs. In particular, there is a range of programs, including cash transfers and credits, that are phased out as individual income increases and various social programs (such as disability insurance and unemployment insurance) in which individuals participate less when labor demand is high (see, for instance, Autor and Duggan 2003; Autor, Dorn, and Hanson 2013; Acemoglu and Restrepo 2020). As a consequence, transfers decline as labor demand rises, and this acts as an additional implicit tax on labor, τ^d. Austin, Glaeser, and Summers (2018) estimate that the public expenditures resulting from a person going into nonemployment was \$4,900 per year between 2010 and 2016 (\$6,300 for those in long-term nonemployment and \$2,300 for the short-term unemployed). This is roughly 8 percent of the average yearly worker compensation during this period, suggesting that social expenditure and disability insurance add an extra 8 percent tax to labor.

II.D. Effective Tax Rates in the United States

Figure 5 in the online appendix depicts the evolution of the average personal income tax and average capital tax rates for C corporations (including both corporate income taxes and personal income taxation) and for S corporations (whose owners only pay personal income taxes and some state-level taxes). Taxes on C corporations' profits decline significantly from 2000 onward, reflecting declines in the statutory corporate income tax rate over time. Taxes on pass-through profits have remained stable around 25 percent, and the average individual income tax has remained close to 15 percent.

24. If we were to use the average payroll tax (about 10 percent in recent years) and the average income tax (about 14.6 percent in recent years), we would end up with a very similar effective tax rate on labor.

Figure 1. Effective Tax Rates on Labor, Software, Equipment, and Nonresidential Structures



Source: Authors' calculations.
Notes: The solid lines depict the observed effective taxes. The dashed lines present the effective taxes that would result if the treatment of allowances had remained as in the year 2000.

Figure 1 presents our estimates for the effective tax rates on labor and different types of capital (in turn computed from effective tax rates on capital and depreciation allowances for C corporations, S corporations, and other pass-through businesses and the differential taxation of capital financed with debt and equity). The solid lines show the effective taxes on software, equipment, nonresidential structures, and labor.

Several points about these effective tax rates are worth noting. First, effective taxes on equipment and software are low compared to the effective taxes on labor. Our benchmark effective tax on labor (which does not include the implicit taxes implied by means-tested programs) hovers around 25.5 percent.[25] In contrast, effective taxes on both equipment capital and software in the 2010s (and before the tax reform of 2017) are around

25. Our estimates imply that the net tax revenue collected by the government with these instruments is roughly 18.6 percent of GDP (25.5 percent × labor income in GDP + 10 percent × net capital income in GDP). This figure matches closely the average share of personal income taxes, corporate taxes, and Social Security contribution in GDP for the period considered in our study (18.7 percent for 1981–2018 in NIPA table 3.1).

10 percent.[26] Second, effective taxes on equipment and software were higher in the 1990s and early 2000s and declined significantly thereafter. This decline is mostly because of the reforms summarized in footnote 16, which have increased depreciation allowances. The dashed lines in figure 1 illustrate the contribution of these reforms by plotting the (counterfactual) effective taxes on different types of capital that would have applied had the treatment of depreciation allowances remained as it was in 2000. They show that about half of the decline in the effective taxes on software and equipment capital is due to the more generous depreciation allowances introduced since 2002. Third, effective taxes on equipment and software decreased further, to about 5 percent, following the 2017 tax reform, which introduced full expensing of these capital expenditures. Finally, because depreciation allowances for structures are lower, the effective tax on nonresidential structures is higher today than tax rates on equipment and software, but in the past the ordering was reversed.

For our purposes, effective tax rates on equipment and software are more relevant, since these are the types of capital that are involved in automation. In what follows, we will summarize the US tax system as an effective tax on labor of $\tau^l = 25.5$ percent and an effective tax on capital of $\tau^k = 10$ percent (the level before the 2017 tax reforms). We will also separately discuss the implications of the reforms in the 2000s and the 2017 tax reform.

III. Does the US Tax Code Favor Automation?

In this section, we investigate whether the US tax system is biased against labor and favors excessive automation. We then explore the implications of different tax reforms.

III.A. Parameter Choices

We first review the estimates of the main parameters in our model. The parameter λ corresponds to the short-run elasticity of substitution between capital and labor. This is the elasticity of substitution between capital and labor holding the amount of automation (and, more generally, the state of technology) constant and without any compositional changes (for example,

26. These effective tax rates are lower than those reported in CBO (2014). Two factors explain the differences. First, and most importantly, the CBO does not incorporate bonus depreciation allowances (based on the argument that these may not be extended in the future). Second, the CBO uses the statutory rate of corporate income tax. As noted above, we do not believe this gives an accurate estimate of the effective tax on capital, since most corporations pay less than the statutory rate.

between firms with different technologies or between industries). Under the assumption that in the short run the allocation of tasks to factors is fixed, this elasticity can be approximated by the short-run elasticity of substitution within establishments, which is estimated to be $\lambda = 0.5$ in Oberfield and Raval (2014).

The other important building block of the production side of our economy is given by the comparative advantage schedules for labor and capital, $\psi^{\ell}(x)$ and $\psi^{k}(x)$. We reduce the dimensions of these functions by assuming that they take isoelastic forms:

$$\frac{\psi^{\ell}(x)}{\psi^{k}(x)} = A \cdot x^{\zeta} \qquad \psi^{\ell}(x) = A \cdot x^{\zeta\upsilon},$$

where $\zeta \geq 0$ controls how the comparative advantage of labor changes across tasks and υ controls the relationship between the comparative and absolute advantage of labor. We take $\upsilon = 1$ as our baseline, which implies that labor is more productive at higher-index tasks (where it has a comparative advantage), while capital has a constant productivity across tasks, as in the "balanced growth" specification in Acemoglu and Restrepo (2018). The online appendix explores the opposite case in which $\upsilon = 0$ and labor is less productive in tasks where it has comparative advantage.

The parameter of comparative advantage ζ (together with λ) shapes the long-run substitution possibilities between capital and labor. In the long run, changes in factor prices will lead to endogenous development and adoption of automation technologies, and as the allocation of tasks to factors changes, there will be greater substitution between capital and labor than implied by λ. The extent of this greater substitution is shaped by the comparative advantage of labor across tasks. In particular, since $\lambda = 0.5$, a lower user cost of capital will increase the labor share of national income in the short run (because capital and labor are gross complements given θ), but as automation adjusts, the labor share could end up lower than it was before the change. Karabarbounis and Neiman (2014) estimate that a 10 percent reduction in the user cost of capital lowers the labor share by 0.83 to 1.67 percentage points in the long run. The midpoint of this range implies $\zeta = 2.12$ in the context of our model.[27]

Turning to labor market imperfections, recall that the wedge ϱ captures the difference between the wage earned by workers and workers'

27. More specifically, Karabarbounis and Neiman (2014) use a constant elasticity of substitution aggregate production function without automation or reallocation of tasks and show that their estimates correspond to a long-run elasticity of substitution in the 1.2–1.5 range.

opportunity cost. This motivates measuring ϱ as the (average) permanent earning loss from job separation. The majority of the estimates of these earning losses in the labor literature are within the range of 5–25 percent with a midpoint of 15 percent.[28] Motivated by this evidence, we choose a baseline value of $\varrho = 0.15$.[29]

The remaining key parameters of our framework are the Hicksian elasticities of labor and capital supply (Hicksian elasticities are the relevant ones in our context because we are focusing on permanent tax reforms). We adopt the following functional forms for utility: $u(\bar{y} - k) = -B \cdot k^{1+1/\varepsilon^k}/(1 + 1/\varepsilon^k) - k$ and $v(\ell) = \ell^{1+1/\varepsilon^\ell}/(1 + 1/\varepsilon^\ell)$, so that the two Hicksian elasticities, $\varepsilon^k \geq 0$ and $\varepsilon^\ell \geq 0$, are constant. The parameters A and B are calibrated to match an aggregate labor share of 56 percent and a net capital share of 26 percent, with the depreciation rate fixed at 5.5 percent per year.

Because our model does not distinguish between the intensive (hours conditional on employment) and extensive (employment) margin, we use the combined elasticity for total hours of work. Chetty and others (2011) report micro elasticity estimates, obtained from differences in tax rates and wages across regions and demographic groups within a country, in the range of 0.46–0.76 (of which 0.33 comes from the intensive margin and 0.13–0.45 comes from the extensive margin). These numbers are close to macro elasticity estimates obtained from tax differences across countries, which are also around 0.7. Because there might be nonlinearities in supply elasticities (see, for example, Mui and Schoefer 2019), and because there is uncertainty about the exact supply elasticities, we explore the implications of labor supply elasticities between 0.46 and 1 in our robustness checks.[30]

28. Couch and Placzek (2010) survey this literature and present their own estimates, suggesting long-run earning declines from separations of 5 percent. Jacobson, LaLonde, and Sullivan (1993) find long-run earning declines of about 25 percent. Davis and von Wachter (2011) report long-run earning losses of 10 percent in normal times and 20 percent in recessions.

29. Some of the earning losses may be due to loss of firm-specific human capital. If productivity gains from firm-specific human capital are shared equally between firms and workers, these would also create a wedge identical in reduced form to our ϱ. We also note that there are other factors that would act like a wedge, generating additional incentives to raise employment. These include negative spillovers from nonemployment on family, friends, and communities and on political behavior (see Austin, Glaeser, and Summers 2018). Because quantifying these effects is more difficult, we are ignoring them in the current paper.

30. In the presence of some types of labor market frictions, the extensive margin changes in employment may take place off the labor supply curve, while intensive margin changes are on the labor supply curve. In table A.3 in the online appendix we show that our main conclusions are robust if we reduce ϱ to 0.075, so that labor market frictions apply only to the extensive margin changes in employment (which make up about half of the variation given the elasticities reported in the text).

The parameter ε^k corresponds to the long-run elasticity with which the supply of capital responds to changes in net returns $d\ln k/d\ln r$ or the keep rate from net capital taxes $d\ln k/d\ln(1 - \tau^k)$ (and is thus different from the demand-side elasticities that are informative about how much investment or capital at the firm level will respond to the user cost of capital). Although there is much uncertainty about this elasticity and many theoretical analyses assume it to be infinite (for example, by imposing time-additive, discounted utility), a number of recent papers estimate it to be much smaller. These studies exploit reforms that change taxes on wealth for different groups of households and find medium-run elasticities that range from 0.2 to 0.65 over four-to-eight-year periods (see Zoutman 2018; Durán-Cabré, Esteller-Moré, and Mas-Montserrat 2019; Jakobsen and others 2020).[31] Using a calibrated life-cycle model and assuming a net after-tax return of $r = 5$ percent, Jakobsen and others (2020) show that their medium-run estimates are consistent with long-run elasticities ranging from 0.58 for the wealthy to 1.15 for the very wealthy. With a lower tax net return of 4 percent (in line with the numbers used in our computation of net effective taxes), long-run capital supply elasticities would be even lower, and conversely, with an after-tax rate of return of 7 percent, these elasticities would range between 1 for wealthy households and 1.9 for the very wealthy (see table 3 in Jakobsen and others 2020). We set our baseline capital supply elasticity to 0.65, which lies at the upper end of the medium-run elasticities reported above and is the average elasticity for the wealthy in Jakobsen and others' (2020) preferred scenario with $r = 5$ percent.[32] We explore the robustness of our results to using a higher elasticity of capital supply in the online appendix.

31. These estimates are from small and fairly open economies, such as Denmark, the Netherlands, and Catalonia, and thus presumably include the response due to the international mobility of capital.

32. We view our choice as conservative given other estimates in the literature. Brülhart and others (2016) estimate the elasticity of capital to after-tax returns using variation across Swiss cantons. They find an elasticity of 1.05 but also show that about a quarter of the effects are driven by migration across cantons and do not involve a change in savings—which is the relevant margin for optimal taxation. In their concluding remarks, they argue that once this response is accounted for, their numbers are comparable to the medium-run estimates of Jakobsen and others (2020). Kleven and Schultz (2014) estimate an elasticity of capital supply with respect to one minus the tax rate on capital income of 0.3, which would imply an even more inelastic response of capital, reinforcing our results. Finally, a related literature finds small elasticities of savings to one minus the estate tax rate, typically about 0.09–0.16 (see Joulfaian 2006; Kopczuk and Slemrod 2000), which also imply less elastic responses of the supply of capital.

III.B. Is the US Tax System Biased against Labor?

We first verify that the US tax system (with $\tau^{\ell} = 25.5$ percent and $\tau^{k} = 10$ percent) is biased against labor. The estimated US taxes comfortably satisfy equation (8) when we use the elasticity estimates in the previous subsection, $\varepsilon^{\ell} = 0.7$ and $\varepsilon^{k} = 0.65$.

Equation (8) implies that current US taxes on labor are too high and US taxes on capital are too low relative to the optimum. In fact, the formulas in proposition 1 for our baseline choice of parameters imply that optimal taxes should be $\tau^{k,r} = 26.65$ percent and $\tau^{\ell,r} = 18.22$ percent, which contrast with the observed taxes of $\tau^{k} = 10$ percent and $\tau^{\ell} = 25.5$ percent. The optimal tax on labor is lower than on capital because the supply elasticities for the two factors are similar, while there is an additional wedge for labor ($\varrho = 0.15$), which the optimal tax system corrects for.

The conclusion that the US tax system is biased against labor is robust to variations in our measurement of effective taxes and our estimates of the elasticities of the supply of capital and labor. The top panel of figure 2 documents that variations in how we compute effective taxes on capital and labor do not change this conclusion. It depicts two contour plots for τ^{ℓ} and τ^{k} that satisfy equation (8) for the baseline values of the remaining parameters ($\varepsilon^{\ell} = 0.7$; $\varepsilon^{k} = 0.65$) and for $\varrho = 0.15$ and $\varrho = 0$. All of our tax estimates lie within these sets and thus comfortably satisfy equation (9) regardless of the value of ϱ.

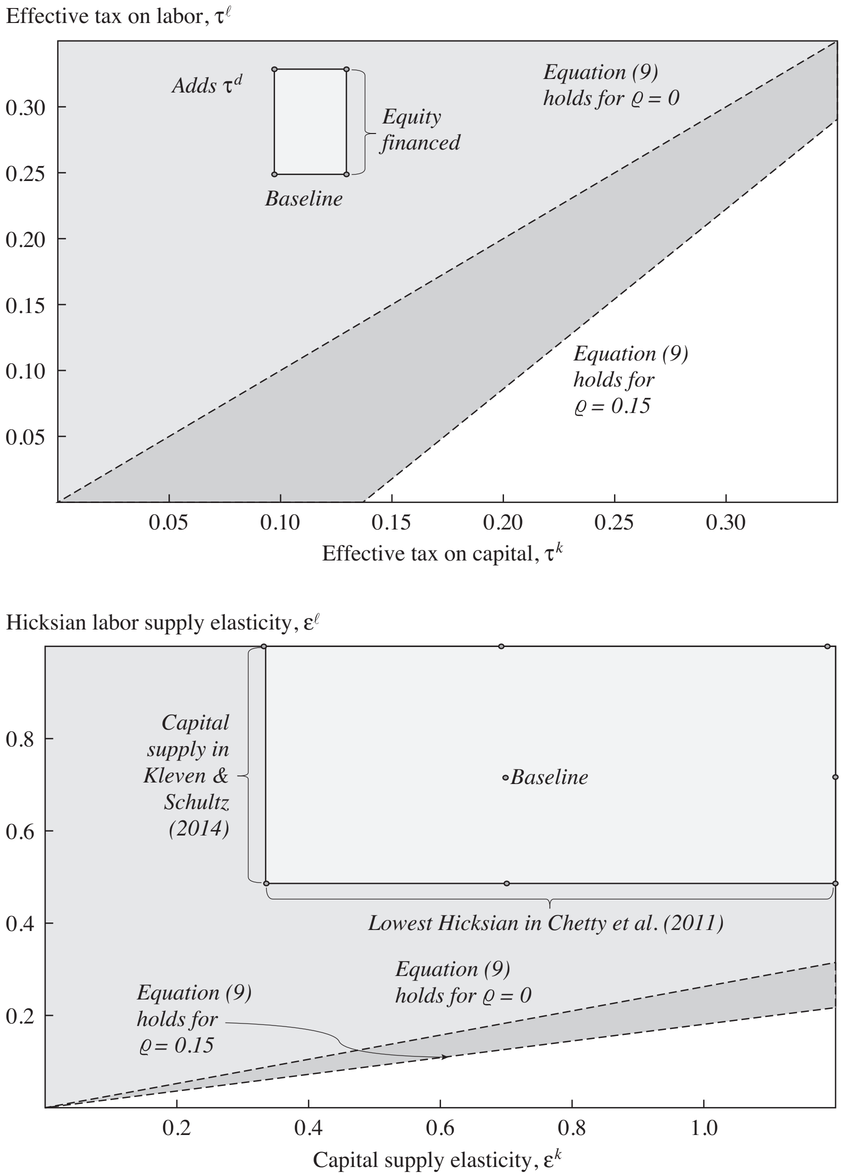
The bottom panel of figure 2 documents that the US tax system remains biased against labor when we vary the elasticities for the supply of capital and labor. The figure presents contour plots for combinations of elasticities ε^{ℓ} and ε^{k} that satisfy equation (9) for our baseline estimates of the US tax system (τ^{ℓ} =25.5 percent;$\tau^{k} = 10$ percent) and again separately for $\varrho = 0.15$ and $\varrho = 0$. We find that even if the capital supply had a unitary elasticity, the US tax system would satisfy equation (9) and would continue to be biased against labor.

III.C. Implications of the US Tax System for Automation and Employment

As discussed in our theory section, the bias against labor in the US tax system will generate excessive automation and lead to lower employment than is socially optimal. We now return to our baseline parameters and investigate the implications of the pro-capital bias of the US tax system for automation, employment, the labor share, and welfare.

As a first step, we compare the implied equilibrium level of automation under the tax system in the 2010s (before the 2017 tax reform),

Figure 2. Contour Plots of Taxes and Elasticities That Verify Equation (9)



Source: Authors' calculations.

Notes: The top panel shows contour plots for estimates of the current US tax system and the bottom panel depicts contour plots for labor and capital supply elasticities to verify the robustness of the claim that the US tax system is biased against labor. Shaded boxes represent the range of estimates we consider in our robustness checks, and in each case we separately mark our baseline estimates. Equation (9) is satisfied for $\varrho = 0$ in the light gray area and for $\varrho = 0.15$ in both the light and the dark gray areas.

Table 1. Equilibrium under the Current Tax System and under Other Potential Scenarios

	Current system (1)	*Ramsey solution (2)*	*Distorting* θ *(3)*	*Distorting* θ *and changing* τ^k *(4)*	*Distorting* θ *and changing* τ^ℓ *(5)*
Tax system					
τ^k (%)	10.00	26.65	10.00	8.39	10.00
τ^ℓ (%)	25.50	18.22	25.50	25.50	24.89
θ	0.276	0.265	0.267	0.265	0.264
τ^A (%)	0.00	0.00	10.15	12.90	13.07
Aggregates (%)					
Employment	—	4.02	1.14	1.59	1.96
Labor share	56.00	56.78	57.93	58.44	58.54
Net output	—	0.44	–0.10	0.16	0.20
C. E. welfare change	—	0.38	0.09	0.14	0.18
Revenue	—	0.00	1.41	0.00	0.00

Source: Authors' calculations.

Notes: This table shows the effective capital and labor taxes, the level of automation, and the automation tax under different scenarios. It also presents the implied changes in employment, output, welfare, and government revenue, and the level of labor share in national income. The first column is for the current US tax system. The second column shows the unconstrained Ramsey solution. Column 3 considers the implications of changing the level of automation, θ via automation taxes (and no other change in policy). Column 4 additionally allows a change in the effective tax on capital, and column 5 considers a change in the effective tax on labor. Change in welfare is in terms of consumption equivalent. See the text for details.

$\tau^\ell = 25.5$ percent and $\tau^k = 10$ percent, to the equilibrium with optimal taxes, $\tau^{\ell,r} = 18.22$ percent and $\tau^{k,r} = 26.65$ percent. Columns 1 and 2 of table 1 present this comparison. Because the optimal tax system encourages the use of labor in production (relative to the US system in the 2010s), it leads to a lower level of automation than currently. Under the optimal tax system, θ declines by 4.1 percent from its equilibrium value in the 2010s.[33] This lower level of automation would also increase the labor share by 0.78 percentage point and, together with the lower labor tax, increase employment by 4.02 percent. Finally, welfare would be higher by 0.38 percent in consumption-equivalent terms (meaning that the welfare gains are equivalent to increasing consumption in period 1 by 0.38 percent). Although this increase in welfare appears small (relative to the change in employment), this is due to the usual intuition related to Harberger's

33. Though the magnitude of a change in θ is not directly interpretable, we can compute the share of employment that would be displaced with the higher level of θ. Given our parameterization of λ, $\psi^\ell(x)$, and $\psi^k(x)$, reducing θ from 0.276 to 0.265 results in 3.3 percent fewer workers displaced by automation.

triangles: because changes in welfare are second order near the optimum, they tend to be smaller than changes in quantities unless we are very far away from this optimum.

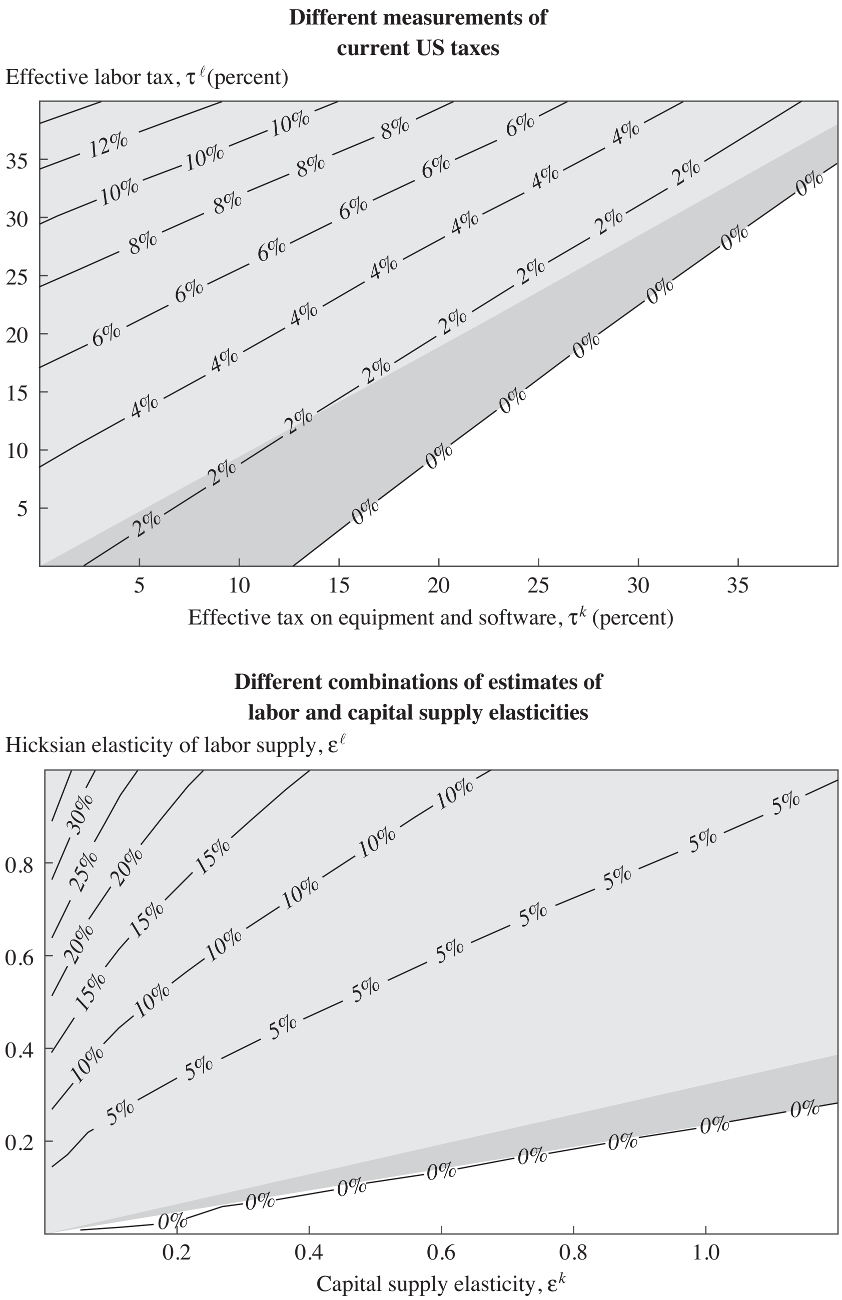
In table 1, we used an effective tax rate on labor of $\tau^\ell = 25.5$ percent, which does not include the additional implicit tax on labor implied by means-tested programs. Table A.4 in the online appendix shows that when we incorporate this additional implicit tax on labor supply and set $\tau^\ell = 33.5$ percent, the employment and welfare gains from changing the current system are amplified. Moving to optimal taxes now increases employment by 6.07 percent, the labor share by 1.09 percentage points, and welfare by 0.81 percent.

The conclusion that we can achieve higher welfare through tax reforms that raise employment and reduce automation is robust in respect to variations in parameters and the measurement of taxes. Figure 3 considers the same range of taxes and parameters as in the two panels of figure 2. The contours in this figure correspond to combinations of current tax rates (top panel) and elasticities (bottom panel) that give the same employment response when we switch from the current tax system to optimal taxes. For a wide range of parameters, optimal taxes induce levels of employment that are 2–10 percent larger than in the current system.

Recall from proposition 2 that when the tax system is biased against labor, the level of automation is not only greater than in the Ramsey solution, but it is also excessively high compared to what would be socially optimal given the tax system. Column 3 in table 1 quantifies this inefficiency by computing the level of automation that would maximize welfare taking the current capital and labor taxes as given.[34] The level of automation that maximizes welfare is $\theta = 0.267$, which is 3.3 percent lower than equilibrium automation. In line with proposition 4, this lower level of automation can be implemented with an automation tax of 10.15 percent, so that a task will be automated only if replacing labor with capital reduces the cost of producing that task by more than 10.15 percent. The automation tax raises employment by 1.14 percent—partially correcting for some of the inefficiencies in the current system and raising welfare—and the

34. The alternative is to follow proposition 2 and maximize the sum of the representative household's utility plus the change in revenue valued at μ (which is the social value of government funds). Here, we simply maximize welfare—given by the representative household's utility—to make the results in this column comparable to the rest of the table. Valuing additional revenues with the multiplier μ leads to higher automation taxes, since reductions in θ have the additional benefit of generating higher labor tax revenue.

Figure 3. Percent Change in Employment Moving from the Current Tax System to the Optimal Tax System



Source: Authors' calculations.

labor share by 1.93 percentage points. Even though equilibrium automation decisions are being distorted, aggregate net output remains essentially unchanged (it declines by 0.10 percent). As already noted, this is because marginal tasks automated under a biased tax system do not increase productivity much (or the automation technology being used in these tasks is so-so).

Column 3 in table 1 allows the planner to change θ, but without modifying the effective tax on capital, τ^k. We next verify that, as implied by proposition 3, if the planner could also modify τ^k (but could not reduce labor taxes), the planner would complement any reform with an automation tax to reduce automation below its market level. This is illustrated in column 4, which shows that in this case the planner achieves higher welfare through a combination of lower capital taxes (τ^k decreases to 8.39 percent) and an automation tax of 12.9 percent, which further reduces θ to 0.265. This alternative tax system would lead to a 2.44 percentage points higher labor share and 1.59 percent more employment.[35]

Finally, column 5 in table 1 turns to a setting where the planner can reduce taxes on labor and distort θ but cannot increase taxes on capital (as mentioned, this scenario may be relevant due to political constraints or fear of capital flight). In this case, the planner would combine a lower labor tax with an automation tax of 13.07 percent, reducing automation again to $\theta = 0.264$ and increasing employment by 1.96 percent and the labor share by 2.54 percentage points.[36]

In summary, our quantitative results show that the current tax system inefficiently favors automation and leads to an employment level that is below the social optimum. The best policy would be to set taxes at their optimal levels, which does not require any further distortions to automation. But if optimal taxes were infeasible, then reducing automation, with or without accompanying changes in other taxes, could reverse some of the inefficiencies in the current tax system and increase

35. As mentioned above, reducing capital taxes may be optimal because the use of capital in tasks in which it has a comparative advantage benefits labor due to complementarity between tasks. In practice, capital might also complement labor in labor-intensive tasks. To capture this possibility, the task production function could be changed to $y(x) = \psi^{\ell}(x) \cdot \ell(x)^{\alpha} \cdot \tilde{k}(x)^{1-\alpha} + \psi^{k}(x) \cdot k(x)$, where $\tilde{k}(x)$ is the capital used to complement labor within tasks. Table A.5 in the online appendix shows that allowing for direct complementarities in this way (with $\alpha = 0.75$) does not change our main findings.

36. Importantly, this can be implemented without raising *any* capital taxes. In particular, a tax on automation can also be implemented via a subsidy to labor of $\tau^A = 13.07$ percent combined with a tax of τ^A on the output of tasks above $\theta^c = 0.264$. This alternative implementation is discussed in proposition A.1 in the online appendix.

employment by 1.14–1.96 percent and the labor share in national income by 1.93–2.54 percentage points.

III.D. Recent Reforms and Effective Stimulus

As described in footnote 16, a series of reforms enacted between 2000 and the mid-2010s significantly reduced effective taxes on equipment and software (from about 20 percent in the year 2000 to about 10 percent). The Tax Cuts and Jobs Act of 2017, which came into effect in 2018, further reduced effective taxes on equipment and software to about 5 percent. These reforms aimed to raise employment by stimulating investment and overall economic activity. In this subsection, we use our calibrated model to study the effectiveness of these reforms and their implications for automation.

Our main finding is that, although all of these reforms increased employment (because they reduced effective taxes), their effects were limited and they increased employment at a large fiscal cost per job created, in large part because they encouraged additional automation. In contrast, we show that alternative reforms reducing labor taxes or combining lower capital taxes with an automation tax could have increased employment by more and at a much lower cost per job.

Column 1 of table 2 reports the market equilibrium for the capital and labor taxes in 2000—$\tau^{\ell} = 25.5$ percent and $\tau^{k} = 20$ percent. Column 2 then documents the impact of the tax cuts on capital enacted between 2000 and the mid-2010s, which reduced the effective tax on software and equipment to 10 percent and reduced government revenue by 10.49 percent. Our model implies that these tax cuts raised employment by a modest 1.01 percent, and did so at a substantial fiscal cost of $162,851 per job. As our theoretical analysis highlights, the lackluster employment response was in part because the lower taxes on capital encouraged greater automation, as shown by the increase in θ. Column 3 turns to the most recent (2017) tax cuts on capital. These are predicted to reduce government revenue by an additional 5.51 percent (or 16 percent relative to the revenue collected in 2000) and encourage further automation, with θ rising to 0.278. The resulting employment gain is again small, 1.47 percent relative to 2000 (or 0.46 percent relative to the mid-2010s), and comes at a fiscal cost per additional job of $169,857.

Columns 4–6 turn to alternative tax reforms that would have cost the same revenue as the capital tax cuts implemented between 2000 and the mid-2010s (10.49 percent of the year 2000 revenue). In column 4, we consider the implications of reducing labor taxes (for example, with a

Table 2. Comparison of Observed Tax Reforms and Reforms Costing the Same Revenue

		Observed reforms		*Alternative reforms*		
	System in 2000 with $\tau^k = 20\%$ (1)	*System in 2010s: reform to $\tau^k = 10\%$ (2)*	*System in 2018: reform to $\tau^k = 5\%$ (3)*	*Labor tax reform (4)*	*Capital tax reform with automation taxation (5)*	*Capital and labor tax reform (6)*
Tax system						
τ^k (%)	20.00	10.00	5.00	20.00	8.58	26.65
τ^ℓ (%)	25.50	25.50	25.50	21.09	25.50	18.22
θ	0.271	0.276	0.278	0.269	0.266	0.265
τ^A (%)	0.00	0.00	0.00	0.00	11.42	0.00
Aggregates						
Employment (%)	—	1.01	1.47	3.56	2.43	5.06
Capital (%)	—	5.64	8.34	1.14	3.07	–2.19
Labor share (%)	56.30	56.00	55.86	56.46	58.15	56.77
Net output (%)	—	2.38	3.49	2.83	2.55	2.83
Cost/revenue per job	—	$162,851	$169,857	$45,954	$67,316	$32,378
Revenue (%)	—	–10.49	–16.00	–10.49	–10.49	–10.49

Source: Authors' calculations.

Notes: This table shows the effective capital and labor taxes and the level of automation for different tax reforms. Column 1 presents the equilibrium under the tax system in the year 2000. Columns 2 and 3 present the resulting changes from the capital tax cuts enacted up to the mid-2010s and then the subsequent capital tax cuts enacted in 2017. Columns 4–6 show the effects of three alternative reforms that would have cost the same as the capital tax cuts enacted between 2000 and the mid-2010s. Column 4 considers cutting the effective labor tax. Column 5 considers a combination of capital tax cuts and a tax to automation. Column 6 considers a combination of lower labor taxes and higher capital taxes.

payroll tax cut) to $\tau^{\ell} = 21.09$ percent and keeping $\tau^{k} = 20$ percent as in 2000. This alternative reform would have increased employment by 3.56 percent and would achieve this at a quarter of the cost of one additional job in column 2. Part of the reason why reducing payroll taxes is much more effective in stimulating employment than cutting capital taxes is that lower payroll taxes reduce automation (θ falls to 0.269) whereas lower capital taxes further increase automation (θ increases from 0.272 to 0.276 between columns 1 and 2).

Column 5 considers another reform, this time combining lower capital taxes with an automation tax (again chosen to cost the same revenue as the tax cuts enacted between 2000 and the mid-2010s). This reform would have also raised employment by more than the reforms of the 2010s, increasing it by 2.43 percent, and would have cost $67,316 per job, which is less than half the cost per job in column 2. Notably, this policy combination involves an even larger tax cut for capital—from 20 percent to 8.58 percent. But crucially, the automation tax simultaneously rolls back any automation that the capital tax cut would have otherwise induced.[37]

Finally, column 6 considers a reform that changes both capital and labor taxes in a welfare maximizing way and costs the same revenue as the reform in column 2. By definition, this reform coincides with the Ramsey solution in column 2 of table 1, and it would have raised the effective capital tax rates to 26.65 percent and reduced the labor tax to 18.22 percent (eliminating the payroll tax almost entirely). We include it in this table to show that, in addition to the 5.06 percent additional increase in employment, such a reform would have had a much smaller cost per job—only $32,378, or about a fifth of the cost per job generated by the capital tax cuts since 2000.

Overall, this discussion shows that, because automation responds to the cost of capital, reducing capital taxes uniformly (via generous depreciation allowances or reductions in corporate taxes) is not an effective way of stimulating employment. Reforms over the last two decades that reduced capital taxes achieved only a modest increase in employment and instead encouraged further automation. Moving forward, reducing payroll taxes or accompanying tax cuts for capital with a tax on automation can more powerfully stimulate economic activity and achieve greater increases in employment at lower fiscal costs.

37. A policy of reducing taxes on capital and at the same time taxing automation is equivalent to lowering the tax on capital by 11.42 percent only for tasks below $\theta = 0.266$. This exceeds the 10 percent tax cut from 2000 to the 2010s. These tax cuts for capital targeted at tasks in which it has a strong comparative advantage allow policymakers to give even larger subsidies to capital accumulation without triggering excessive automation.

III.E. Capital Distortions

Our analysis so far incorporates labor market imperfections via the wedge ϱ but ignores capital distortions. This is motivated by two considerations. First, our starting point is that because of labor market imperfections such as bargaining, search, or efficiency wages, even without any taxes, the level of employment would be too low; the wedge used in our model introduces this property in a simple way. Second, while earning losses from worker displacement provide a natural way of identifying the labor market wedge, there is no simple method for ascertaining whether there are capital wedges and how large they may be.[38] Nevertheless, we have carried out a number of exercises to verify that our conclusions are not unduly affected by this asymmetry in the treatment of capital and labor.

First, if equity finance is not subject to an additional distortion, then the deductions of interest rate payments from taxes in the case of debt finance more than undo any capital market distortions. This is because the interest rate on corporate loans is an upper bound on the capital wedge and is deducted from taxes. Therefore, we can conservatively use effective tax rates on equipment and software that would apply with full equity financing (without any of the reductions in effective capital taxes that come with debt finance). Table A.6 in the online appendix provides analogous results to table 1 in this case. The effective capital taxes are now $\tau^k = 12$ percent, but this has minimal effects on our results. Second, table A.7 in the online appendix repeats our main exercise but now assuming a capital wedge of $\varrho^k = 0.15$—the same as the labor wedge. The employment and welfare gains from moving to optimal taxes are still nontrivial even if about half as large as our baseline estimates. We conclude that our results are not driven by the assumption that there are no capital wedges or the asymmetric treatment of capital and labor.

IV. Extensions

In this section, we discuss two extensions that generalize our model and reinforce our main conclusion that the US tax code favors capital and promotes excessive automation.

38. For example, large corporations that have significant cash at hand should not be using a different external rate of return than their internal rate of return, and their behavior should not be affected by a capital wedge, even if they use external funds. Smaller corporations may face a higher rate of return when borrowing funds, but if investment in these and larger corporations is highly substitutable, this may not correspond to an aggregate capital wedge.

IV.A. Human Capital Investments

The asymmetric treatment of capital and labor may further distort investments in human capital, which may interact with automation decisions. To incorporate this possibility, suppose that the efficiency unit of labor services provided by a worker is augmented by his or her human capital. Assume also that all workers have the same amount of human capital h, so that the efficiency units of labor are now $\ell_h = h \cdot \ell$.[39] The cost of investing in human capital h for ℓ workers is $\dfrac{\ell}{1 + 1/\varepsilon^h} \cdot h^{1+1/\varepsilon^h}$ in terms of the final good of the economy, and $\varepsilon^h > 0$. This parameter will be the elasticity of investment in human capital with respect to changes in wages. Likewise, we take the isoelastic specification of $\nu(\ell)$ used in our quantitative section, so that ε^ℓ is the constant Hicksian elasticity of labor supply.

Incorporating human capital into the labor market-clearing condition, we obtain

$$f_{\ell_h} \cdot (1 - \tau^\ell) \cdot (1 - \varrho) = \ell_h^{1/(\varepsilon^\ell + \varepsilon^h + \varepsilon^\ell \cdot \varepsilon^h)}.$$

The relevant elasticity for the supply of efficiency units of labor has now been replaced by $\varepsilon^\ell + \varepsilon^h + \varepsilon^\ell \cdot \varepsilon^h$, which incorporates the elastic response of human capital and is thus always greater than ε^ℓ. Intuitively, efficiency units of labor can be increased not just by supplying labor but by investing in human capital as well.

The next proposition characterizes optimal taxes in the presence of human capital and shows that labor taxes need to be adjusted to take into account the greater elasticity with which labor services respond to taxation. This pushes in the direction of (relatively) lower labor taxes and, conversely, higher capital taxes.

PROPOSITION 5: Optimal taxes with endogenous human capital

The solution to the Ramsey problem in an environment with human capital satisfies $\theta^r = \theta^m(k, \ell)$ and

$$\frac{\tau^{k,r}}{1 - \tau^{k,r}} = \frac{\mu}{1 + \mu} \frac{1}{\varepsilon^k(k)} \quad \textit{and} \quad \frac{\tau^{\ell,r}}{1 - \tau^{\ell,r}} = \frac{\mu}{1 + \mu} \frac{1}{\varepsilon^\ell + \varepsilon^h + \varepsilon^\ell \cdot \varepsilon^h} - \frac{\varrho}{1 + \mu}.$$

39. This formulation ignores the fact that high-human capital workers may be employed in tasks that are not automated or are complementary to automation technologies. The impact of automation on the employment and wages of different types of workers is explored in Autor, Levy, and Murnane (2003) and Acemoglu and Restrepo (in progress).

Moreover, if an economy has too low a tax on capital and excessive automation without human capital (in the sense of proposition 2), it will a fortiori have too low a tax on capital and excessive automation when there is an elastic response of human capital.

We next provide a back of the envelope quantification of the effect of human capital investments on optimal policy. To do this, we augment our analysis in the previous section with an estimate for the elasticity of human capital, ε^h. We set the elasticity of human capital supply, ε^h, to 0.092. This value is in the mid-range of estimates from the literature on high school completion (Jensen 2010; Kuka, Shenhav, and Shih 2018) and college major choice (Wiswall and Zafar 2015; Beffy, Fougère, and Maurel 2012).[40] This increases the supply elasticity of efficiency units of labor to 0.86, and as a result, the optimal labor tax is now lower, $\tau^\ell = 16.90$ percent, and the optimal capital tax is modestly higher, $\tau^k = 29.21$ percent (see table A.9 in the online appendix). Replacing the current system with optimal taxes leads to more pronounced changes: 1.06 percentage points higher labor share, 5.73 percent increase in employment, and 0.59 percent increase in welfare in consumption-equivalent terms.

IV.B. Endogenous Technology

In our baseline model, increases in θ represent both the development and the adoption of automation technologies. In principle, these two decisions are distinct, even if related. Unless automation technologies are developed, they cannot be adopted. If they are expected to be adopted, then there are greater incentives to develop them. Moreover, as emphasized in Acemoglu and Restrepo (2018), new automation technologies may come at the expense of other technological changes with different implications for capital and labor. For instance, more resources devoted to automation typically imply less effort toward the introduction of new tasks that tend to increase the labor share and demand for labor. If so, a tax structure that favors capital may distort the direction of technological change in a way that disadvantages labor. In this subsection, we provide a simple model to

40. Jensen's (2010) experimental results imply a 0.097 high school completion elasticity in response to perceived returns. Kuka, Shenhav, and Shih (2018) estimate a high school completion elasticity of 0.019–0.086 in response to actual returns and 0.014–0.17 in response to perceived returns. Wiswall and Zafar (2015) estimate elasticities in the range of 0.036–0.062 from the response of college major choice to changes in relative wage premium. Previous estimates in Beffy, Fougère, and Maurel (2012) put the same elasticity in the range of 0.09–0.12. Taken together, these studies imply values for ε^h in the range of 0.014–0.17.

highlight these ideas and show that, with endogenous technology, optimal policy may also need to redirect the direction of technological change, and this is the case even when capital and labor taxes are set optimally.

For brevity, we borrow from the formulation of endogenous technology in Acemoglu (2007, 2010), whereby a (competitive) production sector decides how much capital and labor to use and which technology, from a menu of available technologies, to utilize, while a monopolistically competitive (or simply monopolistic) technology sector decides which menu of technologies to develop and offer to firms.

We consider a menu of technologies consisting of both automation techniques and technologies that increase the productivity and the set of tasks performed by labor, such as the introduction of new tasks considered in Acemoglu and Restrepo (2018). We summarize this menu by Θ with the convention that a higher Θ means a menu that is more biased toward automation technologies. Given menu Θ, firms choose their level of automation θ and their utilization of other technologies ω subject to the feasibility constraint $G(\theta, \omega; \Theta) \leq 0$. Therefore, the index of technologies Θ determines what combinations of automation and other technologies are feasible for final good producers. We denote the production function given θ and ω by $f(k, \ell; \theta, \omega)$, and assume that f_k/f_ℓ is increasing in θ as in our baseline model and decreasing in ω. We further assume that when Θ increases, the set $G(\theta, \omega; \Theta) \leq 0$ includes higher values of θ and lower values of ω, so that a higher Θ enables more adoption of automation technologies and less adoption of other technologies.

The profit-maximizing adoption decision solves the following problem:

$$\{\theta^m(k, \ell; \Theta), \omega^m(k, \ell; \Theta)\} = arg\ max_{G(\theta,\omega;\Theta)\leq 0} f(k, \ell; \theta, \omega).$$

The assumptions on $G(\theta, \omega; \Theta)$ imply that $\omega^m(k, \ell; \Theta)$ is decreasing in Θ and $\theta^m(k, \ell; \Theta)$ is increasing in Θ—so that a higher Θ means a menu of technologies that is more biased toward automation. Hence, as the menu of available technologies becomes more biased toward automation, it crowds out the adoption of nonautomation technologies (such as new tasks or others that increase human productivity).

Finally, we assume that the technology sector charges markups for the use of technologies by final good producers and, via this, captures a constant fraction $\kappa \in (0, 1)$ of the output of these producers (this could be microfounded by assuming that the technology sector sells machines embedding the new technology with a constant markup; see Acemoglu 2007, 2010). We denote the cost of choosing a menu of technologies

Θ by $\Gamma(\Theta)$. Thus, the maximization problem of the technology sector that determines the equilibrium bias of technology is

$$(20) \qquad \max_{\Theta} \kappa \cdot f(k, \ell; \theta^m(k, \ell; \Theta), \omega^m(k, \ell; \Theta)) - \Gamma(\Theta).$$

We make the following assumptions on $\Gamma(\Theta)$:

— $\Gamma(\Theta)$ has a minimum at $\bar{\Theta} \in (0, 1)$. This assumption means that there exists a baseline bias of technology $\bar{\Theta}$, such that deviations from this baseline involve increasing costs. More specifically, deviations from $\bar{\Theta}$ can come in the direction of further automation or further effort devoted to creating new tasks (and thus less automation). Both of these will be more costly than continuing with $\bar{\Theta}$. In the dynamic framework of Acemoglu and Restrepo (2018), $\bar{\Theta}$ corresponds to the state of technology inherited from the past.

— $\Gamma(\Theta)$ is convex, which captures diminishing returns in research directed at changing the bias of technology away from the baseline level $\bar{\Theta}$.

In addition to capital and labor taxes, we allow for subsidies to the use of automation and other technologies in the final good sector to undo the effects of the markup κ and for taxes on the profits of the technology monopolist. Our results do not depend on whether such additional taxes and subsidies exist, but their presence simplifies the expressions and makes them much more closely connected to those in our baseline model in section I.

A market equilibrium satisfies the same market-clearing conditions as in our benchmark economy but is augmented to include the fact that technology adoption decisions of final good producers are given by $\theta^m(k, \ell; \Theta)$ and $\omega^m(k, \ell; \Theta)$, and the equilibrium bias of technology Θ maximizes equation (20). We assume that a market equilibrium exists and is unique and that the solution to equation (20) always involves some interior $\Theta \in (0, 1)$.

We next characterize the solution to the Ramsey problem as in proposition 1. As in our baseline model, we assume that the planner directly controls the development and adoption of technologies (these choices can be implemented with additional taxes as in section I.E).

PROPOSITION 6: Optimal taxes and automation with endogenous technology

The solution to the Ramsey problem with endogenous technology involves capital and labor taxes given as in equation (8) and undistorted adoption decisions (conditional on Θ) given by $\theta^r(k, \ell; \Theta) = \theta^m(k, \ell; \Theta)$ and $\omega^r(k, \ell; \Theta) = \omega^m(k, \ell; \Theta)$. However, if $\Theta^r \gtreqless \bar{\Theta}$, the optimal bias of technology satisfies $\Theta^r \gtreqless \Theta^m$ (i.e., the optimal and market bias of technology are the same if and only if $\Theta^r = \bar{\Theta}$).

The most important implication of this proposition is that, even with optimal taxes on capital and labor, the planner might wish to discourage the development of automation technology. This will be the case when the baseline level of technology is more geared toward automation than what the planner would like to achieve. Put differently, if the economy in question has already gone in the direction of excessively developing automation technologies (which may be a consequence of past distortions or other factors influencing the direction of past technological change), then the planner should intervene by distorting the direction of innovation. The reason for this is that the technology sector does not fully internalize the social surplus created by its technology choices, because of the presence of the term $\kappa < 1$ in equation (20), and thus will not develop the right type of technologies. This result has a close connection to one of the key insights in Acemoglu and others (2012), which established, in the context of optimal climate change policy, that if the economy starts with relatively advanced carbon-emitting, dirty technologies and relatively backward low-carbon, clean technologies, then it is not sufficient to impose Pigovian taxes; rather, optimal policy additionally calls for direct subsidies to the development of clean technologies.[41]

This result has important implications in our context. As our findings in section III suggest, past US tax policy has favored capital and automation. Because these policies have likely led to excessive development of automation technologies, it is not sufficient to simply redress the distortions in the current tax system. Instead, optimal policy may need to intervene to redirect technological change by subsidizing the creation of new tasks and temporarily discourage further effort toward automation innovations at the margin. We leave a quantitative exploration of the implications of endogenous technology development to future work.

V. Concluding Remarks

Automation is transforming labor markets and the structure of work in many economies around the world, not least in the United States. The number of robots in industrial applications and the use of specialized software, artificial intelligence, and several other automation technologies have increased rapidly in the US economy over the last few decades. There has been a concomitant decline in the labor share of national income, wages

41. Note in addition that once the planner can influence the direction of automation technology and set optimal taxes on capital and labor, there is no need to distort the adoption of automation technologies.

have stagnated, and low-skill workers have seen their real wages decline. Many experts believe that these labor market trends are, at least in part, related to automation.

The general intuition among economists (and many policymakers) is that even if automation may have some adverse distributional and employment consequences, policy should not slow down (and certainly not prevent) the adoption of automation technologies because these technologies are contributing to productivity. According to this perspective, policy should instead focus on fiscal redistribution, education, and training to ensure more equally distributed gains and more opportunities for social mobility. But what if automation is excessive from a social point of view?

This paper has argued that the US tax system is likely to be encouraging excessive automation, and if so, reducing the extent of automation (or, more plausibly, slowing down the development and adoption of new automation technologies) may be welfare-improving. We have developed this argument in three steps.

First, we revisited the theory of optimal capital and labor taxation in a task-based framework where there is an explicit decision of firms to automate tasks. We also introduced, albeit in a reduced form, labor market imperfections. Consistent with the classical theory of public finance, if capital and labor taxes are set optimally, automation decisions are optimal in equilibrium. However, away from optimal capital and labor taxes or in the presence of additional constraints on tax decisions, this is no longer the case. Exploiting the structure of our task-based framework, we establish that when the tax system is already biased against labor, it is generally optimal to distort equilibrium automation. The economics of this result is simple but informative: marginal tasks that are automated bring little productivity gains—or in the terminology of Acemoglu and Restrepo (2019a, 2019b), they are "so-so automation technologies"—and as a result, the cost of reducing automation at the margin is second order. When the tax system is biased against labor, the gain from reducing automation and preventing the displacement of labor is first order because it increases employment. In fact, it may even be optimal to reduce automation while at the same time cutting capital taxes (even though the tax system is biased against labor and in favor of capital) because, in contrast to automation, the use of capital in tasks in which capital has a strong comparative advantage is complementary to workers employed in labor-intensive tasks.

Second, we delved into a detailed evaluation of the US tax system in order to map the complex tax code into effective capital and labor taxes. Our numbers suggest that the US system taxes labor heavily and favors capital significantly. While labor is taxed at an effective rate between 25.5 percent and 33.5 percent, capital faces an effective tax rate of about 5 percent (down from 10 percent in the 2010s and 20 percent in the 1990s and early 2000s).

Third, we compared the US tax system to the optimal taxes implied by our theoretical analysis. This exercise confirmed that the US tax system is biased against labor and in favor of capital. As a result, we found that moving from the current US tax system and level of automation to optimal taxation of factors and the optimal level of automation would raise employment by 4.02 percent, the labor share by 0.78 percentage point, and overall welfare by 0.38 percent in consumption-equivalent terms. If optimal taxes can be implemented, there is no need for distorting or taxing automation. If, on the other hand, optimal taxes are infeasible, more modest reforms involving a tax on automation can undo some of the inefficiencies in the current system and increase employment by 1.14–1.96 percent and the labor share by 1.93–2.54 percentage points. In this case, the constrained optimal policy always involves an automation tax in order to discourage the automation of marginal tasks which bring little productivity benefits but significant displacement of labor.

We also showed that a range of realistic generalizations (absent from our baseline framework) reinforce our conclusions and call for even more extensive changes in automation and capital taxation, and under some conditions, it may be optimal to redirect new innovations away from automation.

To simplify the analysis, we focused on an economy with a single type of labor. As noted at the beginning, automation is also associated with increases in inequality (Autor, Levy, and Murnane 2003; Acemoglu and Autor 2011; Acemoglu and Restrepo 2020, in progress). Consequently, slowing down automation may generate additional distributional benefits. These issues are discussed in Guerreiro, Rebelo, and Teles (2017), Thuemmel (2018), and Costinot and Werning (2018). A natural next step is to augment these analyses with the possibility that certain aspects of the tax system may be encouraging excessive automation.

In practice, there are many potential sources of excessive automation. Our objective in this paper has been narrow: to focus on tax reasons for excessive automation. Our companion paper, Acemoglu, Manera, and Restrepo (in progress), shows that even absent tax-related distortions, the market economy tends to generate excessive automation because bargaining power and efficiency wage considerations vary across tasks and this tends to create incentives for firms to automate beyond what is socially beneficial in order to improve their share of rents. Furthermore, as we have already noted, automation-driven job loss may generate negative spillovers on communities and political and social behavior. There may additionally be social factors and corporate strategies concerning the direction of innovation and research (the best minds in many fields being attracted to automation technologies and the most influential companies favoring automation) that further contribute to excessive automation. Quantifying the extent of these other factors is an important area for future research, especially because they have major implications for policy.

Finally, we should briefly comment on how our results relate to two popular policy proposals: wealth taxes and so-called robot taxes. Although our framework suggests that it may be beneficial to increase taxes on capital, wealth taxes on high-wealth individuals may not be the most direct way of achieving this because they would not necessarily increase the effective tax on the use of capital. Increasing corporate income taxes and eliminating or lowering depreciation allowances may be more straightforward ways of implementing higher effective taxes on capital (provided that there are no other distributional or political benefits from wealth taxes). Moreover, our framework highlights that it is not always beneficial to increase taxes on capital: when it is not feasible to implement optimal taxes, reducing automation becomes a central objective (and may even need to be combined with lower taxes on capital). Our automation tax is also different from taxes on robots for the same reasons: it is not a uniform tax on all automation technologies; rather, it is applied to technologies automating tasks above a certain threshold (which are tasks in which humans still have a significant comparative advantage). In fact, our results clarify that instead of taxing all automation technologies, optimal policy often involves subsidizing capital in tasks in which machines have a strong comparative advantage. Last, our analysis also clarifies that if the tax system is reformed so that it is no longer biased against labor and in favor of capital, then employment and welfare can be increased without an automation tax.

ACKNOWLEDGMENTS Prepared for the Brookings Institution Spring Conference of 2020. We are grateful to Alan Auerbach, Jan Eberly, Larry Katz, James Poterba, Ivan Werning, Owen Zidar, and Eric Zwick for excellent comments and suggestions. We gratefully acknowledge financial support from Brookings Institution, Google, the National Science Foundation, Schmidt Sciences, the Smith Richardson Foundation, and the Sloan Foundation.

References

Acemoglu, Daron. 2007. “Equilibrium Bias of Technology.” *Econometrica* 75, no. 5: 1371–409.

Acemoglu, Daron. 2010. “When Does Labor Scarcity Encourage Innovation?” *Journal of Political Economy* 118, no. 6: 1037–78.

Acemoglu, Daron, Philippe Aghion, Leonardo Bursztyn, and David Hemous. 2012. “The Environment and Directed Technical Change.” *American Economic Review* 102, no. 1: 131–66.

Acemoglu, Daron, and David Autor. 2011. “Skills, Tasks and Technologies: Implications for Employment and Earnings.” In *Handbook of Labor Economics*, vol. 4, edited by Orley Ashenfelter and David Card. Amsterdam: Elsevier.

Acemoglu, Daron, David Autor, Joe Hazell, and Pascual Restrepo. 2020. “AI and Jobs: Evidence from Online Vacancies.” Mimeo. Massachusetts Institute of Technology.

Acemoglu, Daron, Andrea Manera, and Pascual Restrepo. In progress. “Automation in Imperfect Labor Markets.” Massachusetts Institute of Technology.

Acemoglu, Daron, and Pascual Restrepo. 2018. “The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment.” *American Economic Review* 108, no. 6: 1488–542.

Acemoglu, Daron, and Pascual Restrepo. 2019a. “Artificial Intelligence, Automation, and Work.” In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb. Chicago: University of Chicago Press.

Acemoglu, Daron, and Pascual Restrepo. 2019b. “Automation and New Tasks: How Technology Displaces and Reinstates Labor.” *Journal of Economic Perspectives* 33, no. 2: 3–30.

Acemoglu, Daron, and Pascual Restrepo. 2020. “Robots and Jobs: Evidence from US Labor Markets.” *Journal of Political Economy* 128, no. 6: 2188–244.

Acemoglu, Daron, and Pascual Restrepo. In progress. “The Labor Share, Displacement and Wage Inequality.” Mimeo. Massachusetts Institute of Technology.

Atkinson, A. B., and J. E. Stiglitz. 1972. “The Structure of Indirect Taxation and Economic Efficiency.” *Journal of Public Economics* 1, no. 1: 97–119.

Atkinson, A. B., and J. E. Stiglitz. 1976. “The Design of Tax Structure: Direct versus Indirect Taxation.” *Journal of Public Economics* 6, no. 1: 55–75.

Auerbach, Alan, and Laurence Kotlikoff. 1987. *Dynamic Fiscal Policy*. Cambridge: Cambridge University Press.

Austin, Benjamin, Edward Glaeser, and Lawrence Summers. 2018. “Jobs for the Heartland: Place-Based Policies in 21st-Century America.” *Brookings Papers on Economic Activity*, Spring, 151–232.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. “The China Syndrome: Local Labor Market Effects of Import Competition in the United States.” *American Economic Review* 103, no. 6: 2121–68.

Autor, David H., and Mark G. Duggan. 2003. "The Rise in the Disability Rolls and the Decline in Unemployment." *Quarterly Journal of Economics* 118, no. 1: 157–206.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118, 4: 1279–333.

Beffy, Magali, Denis Fougère, and Arnaud Maurel. 2012. "Choosing the Field of Study in Postsecondary Education Do Expected Earnings Matter?" *Review of Economics and Statistics* 94, no. 1: 334–47.

Brülhart, Marius, Jonathan Gruber, Matthias Krapf, and Kurt Schmidheiny. 2016. "Taxing Wealth: Evidence from Switzerland." Working Paper 22376. Cambridge, Mass.: National Bureau of Economic Research.

Chamley, Christophe. 1986. "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives." *Econometrica* 54, no. 3: 607–22.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2011. "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins." *American Economic Review* 101, no. 3: 471–75.

Congressional Budget Office (CBO). 2014. "Taxing Capital Income: Effective Marginal Tax Rates Under 2014 Law and Selected Policy Options." Washington. https://www.cbo.gov/sites/default/files/113th-congress-2013–2014/reports/49817–taxingcapitalincome0.pdf.

Cooper, Michael, John McClelland, James Pearce, Richard Prisinzano, Joseph Sullivan, Danny Yagan, Owen Zidar, and Eric Zwick. 2016. "Business in the United States: Who Owns It, and How Much Tax Do They Pay?" In *Tax Policy and the Economy*, vol. 30, edited by Robert A. Moffitt. Chicago: University of Chicago Press.

Costinot, Arnaud, and Iván Werning. 2018. "Robots, Trade, and Luddism: A Sufficient Statistic Approach to Optimal Technology Regulation." Working Paper 25103. Cambridge, Mass.: National Bureau of Economic Research.

Couch, Kenneth A., and Dana W. Placzek. 2010. "Earnings Losses of Displaced Workers Revisited." *American Economic Review* 100, no. 1: 572–89.

Davis, Steven J., and Till von Wachter. 2011. "Recessions and the Costs of Job Loss." *Brookings Papers on Economic Activity*, Fall, 1–72.

Diamond, Peter A., and James A. Mirrlees. 1971. "Optimal Taxation and Public Production I: Production Efficiency." *American Economic Review* 61, no. 1: 8–27.

Durán-Cabré, José M., Alejandro Esteller-Moré, and Mariona Mas-Montserrat. 2019. "Behavioural Responses to the (Re)Introduction of Wealth Taxes: Evidence from Spain." Working Paper 2019/4. Barcelona: Institut d'Economia de Barcelona.

Edgerton, Jesse. 2010. "Investment Incentives and Corporate Tax Asymmetries." *Journal of Public Economics* 94, nos. 11–12: 936–52.

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin. 2013. "The Decline of the U.S. Labor Share." *Brookings Papers on Economic Activity*, Fall, 1–63.

Garrett, Daniel G., Eric Ohrn, and Juan Carlos Suárez Serrato. 2020. "Tax Policy and Local Labor Market Behavior." *American Economic Review: Insights* 2, no. 1: 83–100.

Goldman, Dana, Neeraj Sood, and Arleen Leibowitz. 2005. "Wage and Benefit Changes in Response to Rising Health Insurance Costs." Working Paper 11063. Cambridge, Mass.: National Bureau of Economic Research.

Goolsbee, Austan. 1998. "Investment Tax Incentives, Prices, and the Supply of Capital Goods." *Quarterly Journal of Economics* 113, no. 1: 121–48.

Gruber, Jonathan, and Alan B. Krueger. 1991. "The Incidence of Mandated Employer-Provided Insurance: Lessons from Workers' Compensation Insurance." In *Tax Policy and the Economy*, vol. 5, edited by David F. Bradford. Cambridge, Mass.: MIT Press.

Guerreiro, Joao, Sergio Rebelo, and Pedro Teles. 2017. "Should Robots Be Taxed?" Working Paper 23806. Cambridge, Mass.: National Bureau of Economic Research.

Hall, Robert E., and Dale W. Jorgenson. 1967. "Tax Policy and Investment Behavior." *American Economic Review* 57, no. 3: 391–414.

Hassett, Kevin A., and R. Glenn Hubbard. 2002. "Tax Policy and Business Investment." In *Handbook of Public Economics*, vol. 3, edited by Alan J. Auerbach and Martin Feldstein. Amsterdam: Elsevier.

House, Christopher L., and Matthew D. Shapiro. 2008. "Temporary Investment Tax Incentives: Theory with Evidence from Bonus Depreciation." *American Economic Review* 98, no. 3: 737–68.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. 1993. "Earnings Losses of Displaced Workers." *American Economic Review* 83, no. 4: 685–709.

Jakobsen, Katrine, Kristian Jakobsen, Henrik Kleven, and Gabriel Zucman. 2020. "Wealth Taxation and Wealth Accumulation: Theory and Evidence From Denmark." *Quarterly Journal of Economics* 135, no. 1: 329–88.

Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125, no. 2: 515–48.

Joulfaian, David. 2006. "The Behavioral Response of Wealth Accumulation to Estate Taxation: Time Series Evidence." *National Tax Journal* 59, no. 2: 253–68.

Judd, Kenneth L. 1985. "Redistributive Taxation in a Simple Perfect Foresight Model." *Journal of Public Economics* 28, no. 1: 59–83.

Judd, Kenneth L. 1999. "Optimal Taxation and Spending in General Competitive Growth Models." *Journal of Public Economics* 71, no. 1: 1–26.

Karabarbounis, Loukas, and Brent Neiman. 2014. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129, no. 1: 61–103.

Kleven, Henrik Jacobsen, and Esben Anton Schultz. 2014. "Estimating Taxable Income Responses Using Danish Tax Reforms." *American Economic Journal: Economic Policy* 6, no. 4: 271–301.

Kopczuk, Wojciech, and Joel Slemrod. 2000. “The Impact of the Estate Tax on the Wealth Accumulation and Avoidance Behavior of Donors.” Working Paper 7960. Cambridge, Mass.: National Bureau of Economic Research.

Kuka, Elira, Na’ama Shenhav, and Kevin Shih. 2018. “Do Human Capital Decisions Respond to the Returns to Education? Evidence from DACA.” Working Paper 24315. Cambridge, Mass.: National Bureau of Economic Research.

Lennon, Conor. 2020. “Are the Costs of Employer-Sponsored Health Insurance Passed on to Workers at the Individual Level?” Working Paper. University of Louisville.

Mui, Preston, and Benjamin Schoefer. 2019. “The Labor Supply Curve at the Extensive Margin: A Reservation Wedge Approach.” Paper presented at the Society for Economic Dynamics Annual Meeting, St. Louis, Mo., June 28.

Oberfield, Ezra, and Devesh Raval. 2014. “Micro Data and Macro Technology.” Working Paper 20452. Cambridge, Mass.: National Bureau of Economic Research.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. “Capitalists in the Twenty-First Century.” *Quarterly Journal of Economics* 134, no. 4: 1675–745.

Straub, Ludwig, and Iván Werning. 2020. “Positive Long-Run Capital Taxation: Chamley-Judd Revisited.” *American Economic Review* 110, no. 1: 86–119.

Suárez Serrato, Juan Carlos, and Owen Zidar. 2016. “Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms.” *American Economic Review* 106, no. 9: 2582–624.

Thuemmel, Uwe. 2018. “Optimal Taxation of Robots.” 7317. CESifo Working Paper Series. CESifo Group Munich.

Wiswall, M., and B. Zafar. 2015. “Determinants of College Major Choice: Identification Using an Information Experiment.” *Review of Economic Studies* 82, no. 2: 791–824.

Yagan, Danny. 2015. “Capital Tax Reform and the Real Economy: The Effects of the 2003 Dividend Tax Cut.” *American Economic Review* 105, no. 12: 3531–63.

Zeira, Joseph. 1998. “Workers, Machines, and Economic Growth.” *Quarterly Journal of Economics* 113, no. 4: 1091–117.

Zoutman, Floris T. 2018. “The Elasticity of Taxable Wealth: Evidence from the Netherlands.” Working Paper.

Zwick, Eric, and James Mahon. 2017. “Tax Policy and Heterogeneous Investment Behavior.” *American Economic Review* 107, no. 1: 217–48.

Comments and Discussion

COMMENT BY

LAWRENCE F. KATZ Daron Acemoglu, Andrea Manera, and Pascual Restrepo have reassessed the theory of optimal capital and labor taxation in an elegant task-based framework in which firms make explicit decisions about whether or not to automate tasks. The framework also can account for labor market and capital market imperfections. The authors find that the optimal labor tax rate should be lower than the optimal capital tax rate under (what they argue are) realistic conditions for the US economy of similar effective capital and labor supply elasticities and greater labor market imperfections than capital market imperfections (e.g., a positive labor wedge and no capital wedge). In contrast, the authors carefully document that the actual US effective labor tax rate in the 2010s of 25.5–33.5 percent is much higher than the effective capital tax rate of 5–10 percent for equipment and software. Thus, they conclude that the US tax system is distorted against labor. Furthermore, their framework implies that a tax system biased against labor generates excessive automation with marginal automation being "so-so automation" with only second-order productivity gains but first-order welfare losses from labor displacement. The US tax system's bias against labor thereby leads to lower employment and a lower labor share of national income than optimal in their framework. Finally, the authors argue that if political constraints or international capital flight concerns prevent raising capital taxes to the optimal level, then the second-best policy response is to try to limit automation directly or impose an automation tax to reduce excessive so-so automation and raise employment and welfare.

The authors have produced a provocative, creative, and impressive analysis that certainly makes one rethink how technological advances can potentially harm workers when they are task-replacing automation as opposed to labor-augmenting technological change of the type assumed in standard Solow growth models. And the authors show how a tax system designed (or lobbied) to encourage capital investment can distort automation decisions and lead to too much automation and too little employment. Nevertheless, I do have some concerns about the practical feasibility of their second-best policy of automation restrictions or automation taxes only for capital investment to replace tasks beyond some automation threshold where labor still has a comparative advantage. And I have questions concerning the implications of spillovers across firms or network effects of automation advances as well as how open economy considerations have an impact on the analysis.

In table 1, the authors simulate the potential benefits of task-specific automation taxes that increase capital taxation on automation investments in tasks where labor still has a comparative advantage. But it is not clear how policymakers and tax authorities can identify the task-specific impact of different capital investments. The authors argue that their recommended automation taxes differ from (uniform) robot taxes in only taxing automation above a threshold where it is strongly labor displacing and not taxing capital investments in tasks where machines have a strong comparative advantage. In practice, the US tax system can differentiate taxes on capital assets by asset life and asset categories (structures versus equipment versus software), but it is not obvious how these distinctions translate into different impacts on different types of tasks. More research is likely needed on how investments in different types of capital assets have an impact on different types of worker tasks, such as looking at impacts by occupation and using Bureau of Labor Statistics Occupational Information Network (BLS O*NET) task measures by occupation to determine how differing taxes by capital asset classes might be fine-tuned to more closely approximate an automation tax to discourage marginal (so-so) automation along the lines suggested by the authors. Still, one worries that this will be a daunting task and that even within detailed asset classes (e.g., software) some investments may be labor augmenting and some strongly displacing. And differences in tax rates by asset classes that may be similar to each other based on expected labor displacement effects could generate a whole new range of tax shenanigans through the relabeling of different capital assets to avoid the automation tax or to get favored tax treatment.

Another way to target the automation displacement threshold might be to link capital tax rates to whether a firm keeps employment stable or increases employment as opposed to displacing workers after new investments. But so many difficult-to-measure factors have an impact on employment decisions that again one worries that such a policy could generate unintended distortions. Thus, a more direct and feasible policy approach might be to focus on labor subsidies for workers more at risk of displacement from automation. Such policies could end up looking more like traditional policy responses of wage subsidies for disadvantaged and less-skilled workers or education and training policies to help at-risk workers become more complementary to new technologies.

Open economy considerations also need to be integrated into the authors' framework and policy analysis. Automation that generates only modest productivity gains or cost savings could lead to substantial job displacement in a closed economy setting but could be essential to maintain or expand domestic employment in internationally competitive industries. In fact, Aghion and others (2020) find that for France automation investments significantly increase industry employment (even for less-skilled workers) in industries facing international competition but not in other industries. The implication is that policies that try to limit marginal automation to increase domestic employment could have the opposite effect in the face of foreign competition.

A further difficulty relates to how to identify marginal (or so-so) automation that displaces workers and has only second-order productivity benefits from breakthrough automation investments that could have spillovers to other firms or help create new markets, new tasks, and employment opportunities as emphasized by Acemoglu and Restrepo (2019). For example, one might conclude that autonomous long-haul trucks will end up having only so-so automation, displacing many long-haul truck drivers with only small productivity benefits, and use the tax system to discourage adoption of autonomous vehicles. On the other hand, autonomous long-haul trucks could initially spur new infrastructure and complementary public and private investments that change the nature of interstate highways and linkages of long-haul and short-haul trucking to massively reduce longer-distance domestic transport costs or increase transport speed for products now supplied only locally (e.g., perishable products or artisanal goods), creating new markets, more short-haul trucking and logistics jobs, and overall improved employment activities. Public infrastructure investments, R & D policies, education and training policies, and technology extension policies may alter the impacts of automation such that so-so automation for individual firms that appears labor displacing might be beneficial for workers in the broader economy. The more positive impacts of automation on productivity growth and worker outcomes in the mid-twentieth century of shared prosperity and stable labor share as opposed to the twenty-first century of declining labor share and weak wage growth might have more to do with such complementary policies than changes in tax treatment of capital versus labor.

Finally, the authors' conclusion that optimal labor taxes should be lower than optimal capital taxes and that the US tax system is distorted against labor, in the sense of their equation (9), depends on the capital and labor supply elasticities being of similar magnitude or of labor market distortions being larger than capital market distortions. The authors admirably present a lot of robustness checks to different values for the labor and capital supply elasticities and the size of labor and capital market distortions. But I do remain concerned that the large international mobility of capital means that the effective capital supply elasticity to US taxes could be in the high range of their estimates (even greater than one) and that effective labor supply elasticities could be below the range they consider. Further disaggregation across education, age, and gender groups in labor markets with different labor supply elasticities could help sharpen the analysis. And the large decline in the worker power in the United States documented by Stansbury and Summers (2020) over recent decades implies a substantial reduction in the labor wedge that could be of similar magnitude to (and thereby offset) the decline in capital taxes versus labor taxes and should have operated to ameliorate the problem of excessive automation in the authors' framework from tax distortions.

REFERENCES FOR THE KATZ COMMENT

Acemoglu, Daron, and Pascual Restrepo. 2019. "Automation and New Tasks: How Technology Displaces and Reinstates Labor." *Journal of Economic Perspectives* 33, no. 2: 3–30.

Aghion, Philippe, Céline Antonin, Simon Bunel, and Xavier Jaravel. 2020. "What Are the Labor Market and Product Market Effects of Automation? New Evidence from France." Working Paper. https://www.xavierjaravel.com/papers.

Stansbury, Anna, and Lawrence H. Summers. 2020. "The Declining Worker Power Hypothesis: An Explanation for the Recent Evolution of the American Economy." In the present volume of *Brookings Papers on Economic Activity*.

COMMENT BY

ERIC ZWICK Acemoglu, Manera, and Restrepo provide a thought-provoking perspective on capital taxes and automation. The paper is clearly written and the model's logic is intuitive. I especially appreciate their careful attention to detail in mapping the model onto its empirical analogues in the US tax system. My comments focus on capital tax questions and quibbles, both in mapping the model to the data and in applying the results to inform capital tax policy. To me, the question—Where does automation rank relative to other considerations in evaluating tax policy?—remains open. But I suspect we have not heard the last from this team. In the words of a famous robot from popular culture, "[They'll] be back!"

EVALUATING KEY CAPITAL MARKETS ASSUMPTIONS The paper's primary contribution is a set of theoretical results on the potential for welfare-improving taxation of automated tasks or complementary capital. These results are then calibrated under certain assumptions to allow quantitative statements (1) on whether the tax code suboptimally favors automation and (2) on what optimal tax rates on labor and capital should be. Under the offered set of assumptions, the model's results hold. My first comment concerns which assumptions are important for the main results and whether these assumptions are suitably chosen. I focus on three parameters of interest: the elasticity of capital supply, the capital markets wedge, and the effective capital tax rate.

First, the model's results accord with a Ramsey rule intuition that optimal taxes are inversely proportional to the elasticities of capital and labor supply. Thus, the relative elasticity of capital versus labor is crucial for the quantitative exercise. The authors' baseline parameterization sets the capital supply elasticity to 0.65, slightly below the labor supply elasticity of 0.7. The source for this assumption is recent quasi-experimental research on wealth and savings taxes, for example, the recent paper by Jakobsen and others (2020) which studies a recent wealth tax reform in Sweden.

I believe the conventional wisdom on labor versus capital remains that capital supply is more elastic than labor supply, so I was surprised to see them calibrated to be about the same. Taking estimates from the wealth tax literature to this setting requires a nontrivial extrapolation, namely, that the local effects estimated for a small subpopulation can inform the aggregate capital supply elasticity.

Moreover, the wealth and savings tax literature are not the only useful sources for such estimates. For example, House and Shapiro (2008) use the first round of bonus depreciation incentives to estimate the elasticity of capital supply, subject to the assumption that demand elasticities for long-lived goods in response to temporary subsidies are infinite. Their capital supply elasticities range from 6 to 14, an order of magnitude larger than the assumed elasticity here. Of course, the tradition in theoretical corporate tax incidence going back to Harberger (1962) and Feldstein (1974) has been to permit the capital supply elasticity to approach infinity in the medium and long runs. While I believe finite capital supply elasticities are well justified and supported in the data, the bottom line is that the working assumption here appears to me to be nonstandard, potentially controversial, and quantitatively relevant.

Second, the core theoretical results follow from the interaction between automation and the assumption of an unrelated labor market wedge that implies equilibrium employment is suboptimally low. Key here is that there is not a similar capital market wedge. Of course, there is a large body of research on capital market wedges, deriving from information asymmetries, agency problems, imperfect contractibility, and so on. An influential line of *Brookings Papers* going back to Fazzari, Hubbard, and Petersen (1988), along with the entire field of corporate finance, seems to invite us to consider relaxing the perfect capital markets assumption. It might be worth generalizing the results to a setting with a capital market wedge and making statements a function of the relative wedges. I suspect some of the results on the interaction between capital taxes and automation policies depend on the assumption of no capital market frictions.

Third, the authors make a tremendous and laudable attempt to calibrate the model to match the recent history and current level of effective labor and capital taxes. In modeling capital taxes, they account for the multitudinous idiosyncrasies of the American tax system, including depreciation incentives, business entity taxes across all corporate forms, payout taxes, effective taxes on debt versus equity finance, and differences across structures, equipment, and software investment. I recommend that any interested reader spend time in the paper's tax appendix, which details these calculations and will be helpful to others working in this area.

In the baseline calibration, the paper calculates an effective tax rate on capital in the 2010s of approximately 10 percent. I have one concern about this calculation, which pertains to the temporary nature of bonus depreciation. The main empirical inputs into these effective rates are National Income and Product Accounts (NIPA) aggregates for corporate tax revenues relative to gross operating surplus. These measures are poorly suited to capture the dynamic effects of bonus depreciation on effective tax rates. Tax revenues fall temporarily in stimulus years, but at the mechanical expense of higher taxes in the future. I worry that the current calculation of effective rates does not fully account for this dynamic. Nevertheless, it is relatively uncontroversial to claim that the tax burden on capital has fallen over time, which is the more important message of the paper.

My overall takeaway on these capital market assumptions is that defensible changes to them would quantitatively alter the authors' results. At the same time, the qualitative results are on firmer ground. Accounting for automation in considering tax policy changes could well be an important consideration going forward.

THE ROLE OF BONUS DEPRECIATION When I was on the "job market" presenting the results from Zwick and Mahon (2017), I had to work pretty hard to convince people that bonus depreciation had any effect on anything. Times have changed. The logic in the current paper implies that, not only does bonus depreciation matter for capital accumulation, it also materially biases the factor mix of production away from labor and toward automation.

How much does bonus depreciation distort investment and disfavor labor? Bonus depreciation accelerates the timing of deductions but does not change their amount. Thus, in contrast to changes in the corporate tax rate or an investment tax credit, its value to firms is driven only by discounting.[1] When interest rates are low, as they have been for the last decade, the effective subsidy is relatively small.[2] Accordingly, the aggregate effective tax rate shouldn't be very sensitive to bonus depreciation incentives. As a result, we likely need a very large substitution elasticity between labor and automation for bonus depreciation to be quantitatively relevant for aggregate automation trends.

This logic is also why the Joint Committee on Taxation (JCT) and the Treasury Department do not score bonus depreciation incentives as being very expensive in the ten-year budget window.[3] As noted above, this logic is also somewhat at odds with the authors' current approach to measuring the impact of bonus depreciation on effective tax rates without accounting for future tax payments.

1. My reading of the 1980s tax history is that the investment tax credit in the 1981–1986 period, when combined with the accelerated cost recovery system (ACRS) and a more generous treatment of passive losses, was more generous to capital than recent changes to accelerated depreciation for equipment expenditures. Relative to this period, I am not sure that current capital taxes are significantly more favorable.

2. Zwick and Mahon (2017) argue this may not be the case for firms facing financial frictions.

3. For example, in the JCT score for the Tax Cuts and Jobs Act, the ten-year revenue estimate for expanded depreciation incentives shows an increase in revenue in later years, even before the policy sunsets. Over the ten-year window, the cost of depreciation changes is less than 10 percent of the cost of the corporate tax rate cut. See, for example, JCX-67-17, "Estimated Budget Effects of the Conference Agreement for H.R.1."

15616-04b-Acemoglu_Com&GD_3rdPgs.indd 292 11/19/20 9:34 AM

It would be terrific to see additional empirical work evaluating the effects of bonus depreciation on labor markets. To date, we have seen work documenting that investment increases overall, and more so for firms valuing liquidity and immediate benefits (House and Shapiro 2008; Zwick and Mahon 2017). One labor market effect of greater investment demand is through output effects among capital suppliers, where we would expect to see *higher* employment.

Zwick and Mahon (2017) also present evidence that the wage bill increases at the firm level among firms buying more equipment due to bonus. This result has been confirmed at the local labor market level by Garrett, Ohrn, and Suárez Serrato (2020) and Ohrn (2019). These two papers also show that employment either remains unchanged or increases. In a fascinating recent study, Tuzel and Zhang (2019) find that, among firms buying more equipment in response to depreciation incentives, skilled labor increases while unskilled labor falls.

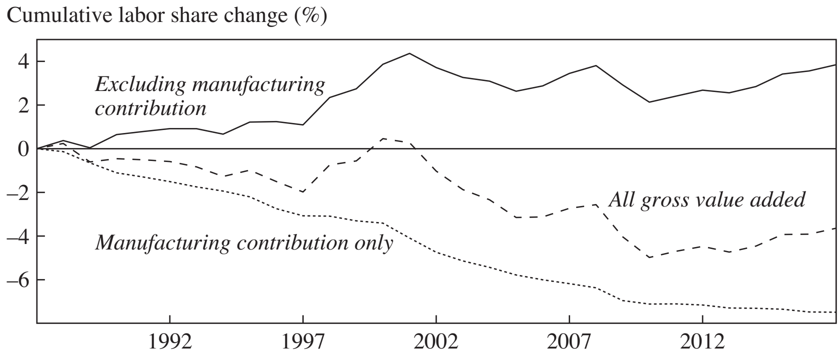
Taken together, the existing evidence points toward potential complementarity between labor and capital demand induced by bonus depreciation. In the terminology of Acemoglu and Restrepo (2018), it is possible the productivity effect dominates the displacement effect. More work is needed to establish the robustness of these results and to investigate the extent to which such investment incentives promote automation.

A TASK-SPECIFIC TAX IN PRACTICE? The authors highlight the theoretical appeal of an *automation tax*, described as "an additional tax on the use of capital in tasks where labor has a comparative advantage." The intuition for this result is clear: such a tax has a first-order benefit in increasing labor demand away from its inefficiently low level and only a second-order cost because so-so automation is only so-so. But what are the practical policy implications?

Perhaps I have read too many papers by Joel Slemrod, but I couldn't help imagining the various strategies that firms and their consultants could devise to avoid such a tax. The literature on automation has attempted to identify those occupations that face automation risk, but we are very far from being able to codify such a system into policy. Were we to take on such a challenge, we would surely face the implications of what I call the *Slemrod conjecture* (Slemrod and Kopczuk 2002): tax avoidance is weakly increasing in the number of tax instruments.

This risk has been realized in the case of bonus depreciation, arguably a much easier policy to implement and enforce, as it builds upon preexisting rules. Subsequent to bonus depreciation's enactment, a new consulting service called "cost segregation" has grown in popularity. These consultants

Figure 1. Cumulative Labor Share Decline with and without Manufacturing



Source: Smith and others (2019, fig. 5); data from US Bureau of Economic Analysis (BEA).

advise companies on how to adjust their accounting to relabel ineligible investment expenditures (for example, fixed internal features of new buildings) as shorter-lived expenditures to be depreciated under bonus. This industry has even spawned an organization of cost segregation experts, the American Society of Cost Segregation Professionals, who have developed standards, educational material, and even a code of ethics. While we can debate the likely employment effects of an automation tax for workers, I am more confident (and also concerned) that such a tax will help promote the full employment of accountants.

REVISITING THE LABOR SHARE'S DECLINE My final comment concerns the authors' broader motivating question, which has animated research in macroeconomics over the last five to ten years: What is driving the decline in the labor share? A more specific version of this question concerns the role of tax policy in the labor share's decline.

As a starting point, let us remind ourselves that manufacturing is the most quantitatively important sector of the economy for understanding the labor share's trend since the 1980s. This fact is reasonably well known but sometimes underemphasized (figure 1).[4] Less well appreciated is that

4. For example, both Karabarbounis and Neiman (2014) and Autor and others (2020) emphasize the broad-based nature of declines in the labor share. See Charles, Hurst, and Schwartz (2019) for a recent survey with new facts.

this secular decline in the manufacturing sector's contribution to the labor share was offset—fully until 2000 and partly since then—by a rise in the contribution from services. To see this, it is important to recognize evolving tax incentives to characterize owner-manager payments as labor versus profits in the skilled service sector (Smith and others 2019).

These facts are useful for the automation story, because manufacturing is one of the sectors most exposed to the rise of robots and other process automation (Acemoglu and Restrepo 2020). Perhaps automation induced by recent changes in tax policy is an important driver of labor share declines, especially in manufacturing?

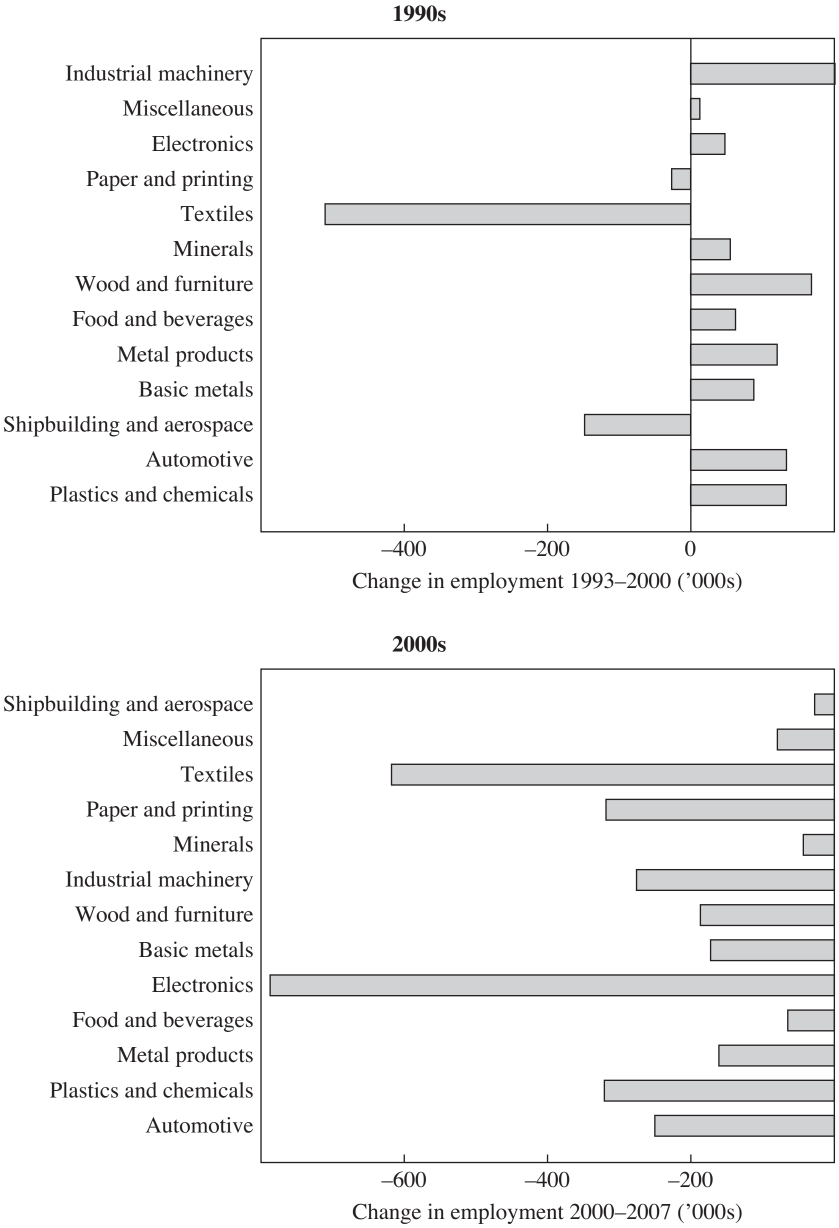
Figure 2 takes a closer look at the trends in employment within manufacturing over the two time periods of study in Acemoglu and Restrepo (2020), where we have sorted the industries in each time period from lowest to highest in their exposure to automation.[5]

In the 1993–2000 period, only textile manufacturing and aerospace manufacturing contribute to employment declines, with textiles accounting for most of the decline. In the 2000–2007 period, the decline in manufacturing employment is broad-based and especially large in electronics but also continues for textiles. Notably, the within-manufacturing correlation between employment declines and automation exposure is present in the later period, consistent with Acemoglu and Restrepo (2020), but not especially pronounced.

In contrast, the evidence on the aggregate role of trade exposure and offshoring for the decline in employment and the labor share is well established and quite strong (Elsby, Hobijn, and Şahin 2013; Autor, Dorn, and Hanson 2013; Pierce and Schott 2016). An open question concerns whether tax incentives amplify these forces. Federal tax policy up to and including the most recent round of tax reforms features strong incentives to locate both profits and real activity offshore. One could argue that bonus depreciation, by prioritizing capital expenditures within the United States, leans against these incentives. I would love to see more research in this area.

5. While it would be ideal to look at contributions to the aggregate labor share directly, mapping industries to the broader US Bureau of Economic Analysis (BEA) sectors is nontrivial. Employment declines likely offer a useful and policy-relevant view into understanding labor share trends.

Figure 2. Employment Declines Sorted by Automation Exposure



Source: Author's calculations combining manufacturing employment statistics from the Census with automation exposure from Acemoglu and Restrepo (2020).

Note: The industries are sorted from lowest to highest in terms of automation exposure. This measure differs across the early and later periods, hence the difference in industry ordering.

REFERENCES FOR THE ZWICK COMMENT

Acemoglu, Daron, and Pascual Restrepo. 2018. "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review* 108, no. 6: 1488–542.

Acemoglu, Daron, and Pascual Restrepo. 2020. "Robots and Jobs: Evidence from US Labor Markets." *Journal of Political Economy* 128, no. 6: 2188–244.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103, no. 6: 2121–68.

Autor, David H., David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics* 135, no. 2: 645–709.

Charles, Kerwin Kofi, Erik Hurst, and Mariel Schwartz. 2019. "The Transformation of Manufacturing and the Decline in US Employment." *NBER Macroeconomics Annual* 33:307–72.

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin. 2013. "The Decline of the U.S. Labor Share." *Brookings Papers on Economic Activity*, Fall, 1–63.

Fazzari, Steven M., R. Glenn Hubbard, and Bruce C. Petersen. 1988. "Financing Constraints and Corporate Investment." *Brookings Papers on Economic Activity*, no. 1: 141–206.

Feldstein, Martin. 1974. "Incidence of a Capital Income Tax in a Growing Economy with Variable Savings Rates." *Review of Economic Studies* 41, no. 4: 505–13.

Garrett, Daniel G., Eric Ohrn, and Juan Carlos Suárez Serrato. 2020. "Tax Policy and Local Labor Market Behavior." *American Economic Review: Insights* 2, no. 1: 83–100.

Harberger, Arnold C. 1962. "The Incidence of the Corporation Income Tax." *Journal of Political Economy* 70, no. 3: 215–40.

House, Christopher L., and Matthew D. Shapiro. 2008. "Temporary Investment Tax Incentives: Theory with Evidence from Bonus Depreciation." *American Economic Review* 98, no. 3: 737–68.

Jakobsen, Katrine, Kristian Jakobsen, Henrik Kleven, and Gabriel Zucman. 2020. "Wealth Taxation and Wealth Accumulation: Theory and Evidence from Denmark." *Quarterly Journal of Economics* 135, no. 1: 329–88.

Karabarbounis, Loukas, and Brent Neiman. 2014. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129, no. 1: 61–103.

Ohrn, Eric. 2019. "The Effect of Tax Incentives on US Manufacturing: Evidence from State Accelerated Depreciation Policies." *Journal of Public Economics* 180:1–14.

Pierce, Justin R., and Peter K. Schott. 2016. "The Surprisingly Swift Decline of US Manufacturing Employment." *American Economic Review* 106, no. 7: 1632–62.

Slemrod, Joel, and Wojciech Kopczuk. 2002. "The Optimal Elasticity of Taxable Income." *Journal of Public Economics* 84, no. 1: 91–112.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. "The Rise of Pass-Throughs and the Decline of the Labor Share." Working Paper. http://ericzwick.com/labshare/labshare.pdf.

Tuzel, Selale, and Miao Ben Zhang. 2019. "Economic Stimulus at the Expense of Routine-Task Jobs." Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3072733.

Zwick, Eric, and James Mahon. 2017. "Tax Policy and Heterogeneous Investment Behavior." *American Economic Review* 107, no. 1: 217–48.

GENERAL DISCUSSION Olivier Blanchard was disappointed that there was not more direct evidence on the effect of relative taxes on automation. He expected that there would be differences in the relative tax rates across countries, sectors, and distinct types of investments, which would lead countries or sectors to choose different technologies. Blanchard also wondered whether the reduced form labor supply specification used in the paper accounted for the fact that some low-skilled workers are going to be unemployable, because of the minimum wage effect. In addition to low-skilled workers, some skilled workers are going to be replaced by machines, and if their skills are not transferrable, there will be a large loss in income.

Robert Hall found the observations made in the paper striking and carefully thought out. He noted that the conclusions made by the authors stemmed from the parameters chosen for the capital supply elasticity. In the early 1970s, Robert Barro had a convincing argument that the capital supply elasticity was infinite.[1] In that case, Hall suggested that the optimal capital tax rate would be around 10 percent compared to the 27 percent optimal capital tax suggested by the authors. Hall also expressed related concerns about the authors' decision to use 0.65 as the capital supply elasticity. He thought that a capital supply elasticity that is less than one is extremely low and should not be taken as a fact. This is because capital in the United States and other wealthy economies is supplied by wealthy individuals, thus the behavior of the typical consumer is not relevant to the question of the capital supply because they are typically not marginal participants in the capital market. He recognized that there are obstacles to measuring capital supply elasticity but suggested that accounting for extreme inequality and uneven distribution of wealth will have a big effect on correctly measuring the capital supply elasticity.

1. Robert J. Barro, "Are Government Bonds Net Wealth?," *Journal of Political Economy* 82, no. 6 (1974), 1095–117.

James Poterba observed that it would be helpful to describe how far the United States tax system deviated from one that would deliver an undistorted "production efficiency" outcome, a benchmark raised in work by Diamond and Mirrlees.[2] He pointed out that there are other potential externalities that the authors might want to consider. For example, some argue that equipment should receive more favorable tax treatment than structures because there are some positive externalities associated with equipment investments, such as the acceleration of technology deployment. Also, low tax rates on software investments might in part serve as a response to imperfections in the intellectual property market, such as limitations on patenting software. Poterba did not have a way to attach numbers to such considerations or to other capital market imperfections, but he believed that quantitative analysis of these factors would result in a much richer discussion of the tax burdens on capital and labor.

Daron Acemoglu agreed that the elasticity of capital supply is an important parameter and that there was a great deal of uncertainty around it. However, he stated that capital may be very elastic when we focus on firms, but the aggregate supply coming from the consumer side may not be as elastic. Regarding Robert Hall's comment, Acemoglu noted that infinite elasticity of capital is a natural benchmark that economists gravitate toward, but models that assume infinite elasticity of capital supply are extremely special and do not reflect the reality. Further, evidence does not support an infinite elasticity of capital. Acemoglu stated that the best papers, from his point of view, found that the elasticity of capital supply was around 0.7, but this is a noisy area in the literature. He agreed with James Poterba that this is an area where further discussion is needed.

Acemoglu further argued that a high or infinite supply elasticity would not have an impact on employment. Thus, the large effects on employment shown in the paper were a result of the supply elasticity parameters that Acemoglu and his coauthors chose. He noted that evidence suggests that capital is not as elastic as economists normally presume. Citing the work of Diamond and Mirrlees mentioned earlier, Acemoglu argued that although the theorem suggested was correct, it only worked under extremely specific assumptions. The theorems do not apply in the absence of the extremely specific assumptions. Acemoglu agreed that the practicality of an automation tax needs to be rethought. However, he suggested that implementing the Ramsey taxes would eliminate the need

2. Peter A. Diamond and James A. Mirrlees, "Optimal Taxation and Public Production I: Production Efficiency," *American Economic Review* 61, no. 1 (1971), 8–27.

for automation taxes. If capital is excessively subsidized as shown in the paper, and if implementing an automation tax is impractical, then reducing the excessive subsidy to capital and reducing payroll taxes would be welfare enhancing.

Pascual Restrepo clarified that they had conducted several robustness checks with different values of capital elasticity. The results on welfare and employment gain are robust, even for higher values of capital elasticity.

MARCO DEL NEGRO
Federal Reserve Bank of New York

GIORGIO E. PRIMICERI
Northwestern University

MICHELE LENZA
European Central Bank

ANDREA TAMBALOTTI
Federal Reserve Bank of New York

What's Up with the Phillips Curve?

ABSTRACT The business cycle is alive and well, and real variables respond to it more or less as they always did. Witness the Great Recession. Inflation, in contrast, has gone quiescent. This paper studies the sources of this disconnect using vector autoregressions and an estimated dynamic stochastic general equilibrium model. It finds that the disconnect is due primarily to the muted reaction of inflation to cost pressures, regardless of how they are measured—a flat aggregate supply curve. A shift in policy toward more forceful inflation stabilization also appears to have played some role by reducing the impact of demand shocks on the real economy. The evidence rules out stories centered around changes in the structure of the labor market or in how we should measure its tightness.

The recent history of inflation and unemployment is a puzzle. The unemployment rate has gone from below 5 percent in 2006–2007 to 10 percent at the end of 2009, and back down below 4 percent in 2018–2019. These fluctuations are as wide as any experienced by the US economy in the postwar period. In contrast, inflation has been as stable as

Conflict of Interest Disclosure: Marco Del Negro is vice president in the Macroeconomic and Monetary Studies Function of the Research and Statistics Group of the Federal Reserve Bank of New York and coeditor of the *Journal of Applied Econometrics*; Michele Lenza is head of section for the Monetary Policy Research Division of the European Central Bank, the Forecasting and Business Cycle group leader and research coordinator for the European Central Bank, and a member of the Forecast Steering Committee for the European Central Bank; Giorgio E. Primiceri is a professor of economics at Northwestern University and coeditor of the *American Economic Journal: Macroeconomics*; Andrea Tambalotti is vice president of the Federal Reserve Bank of New York. Beyond these affiliations, 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 views expressed in this paper are those of the authors and do not necessarily reflect those of Northwestern University, the *American Economic Journal*, the *Journal of Applied Econometrics*, the Federal Reserve Bank of New York, or the European Central Bank.

ever, with core inflation almost always between 1 and 2.5 percent, except for short bouts below 1 percent in the darkest hours of the Great Recession.

Much has been written about this disconnect between inflation and unemployment. In the early phase of the expansion, when unemployment was close to 10 percent and inflation barely dipped below 1 percent, the search was for the "missing deflation" (Hall 2011; Ball and Mazumder 2011; Coibion and Gorodnichenko 2015; Del Negro, Giannoni, and Schorfheide 2015; Lindé and Trabandt 2019). More recently, with unemployment below 5 percent for almost four years and inflation persistently under 2 percent, attention has turned to factors that may explain why inflation is not coming back (Powell 2019; Yellen 2019). Beyond this recent episode, a reduction in the cyclical correlation between inflation and real activity has been evident at least since the 1990s (Atkeson and Ohanian 2001; Stock and Watson 2007, 2008, 2019; Zhang, Chan, and Cross forthcoming). The literature, which we review in more detail below, has focused on four main classes of explanations for this puzzle: (1) mismeasurement of either inflation or economic slack; (2) a flatter wage Phillips curve; (3) a flatter price Phillips curve; and (4) a flatter aggregate demand relationship, induced by an improvement in the ability of policy to stabilize inflation.

This paper tries to distinguish among these four competing hypotheses using a variety of time series methods. We find overwhelming evidence in favor of a flatter price Phillips curve. Some of the evidence is also consistent with a change in policy that has led to a flatter aggregate demand relationship.

The analysis starts by illustrating a set of empirical facts regarding the dynamics of inflation in relation to other macroeconomic variables, using a vector autoregression (VAR). Many of these facts are already known, but the dynamic, multivariate perspective offered by the VAR makes it easier to consider them jointly, enhancing our ability to point toward promising explanations for the phenomenon of interest. First, goods inflation has become much less sensitive to the business cycle since 1990, consistent with most of the literature on the severe illness of the Phillips curve. Second, this is true to a lesser extent for nominal wage inflation: the wage Phillips curve is in better health than its price counterpart, as also found by Coibion, Gorodnichenko, and Koustas (2013), Coibion and Gorodnichenko (2015), Galí and Gambetti (2019), Hooper, Mishkin, and Sufi (2020), and Rognlie (2019). Third, there is little change before and after 1990 in the business cycle dynamics of the most popular indicators of inflationary pressures relative to each other, especially when compared to the pronounced reduction in the responsiveness of inflation.

These indicators include measures of labor market activity, such as the unemployment rate and its deviations from the natural rate, hours, the employment-to-population ratio, and unit labor costs, as well as broader notions of resource utilization, such as GDP and its deviation from measures of potential. Fourth, the decline in the sensitivity of inflation to the business cycle is heterogenous across goods and services. In particular, Stock and Watson (2019) document that the relationship between cyclical unemployment and inflation has changed very little over time for certain categories of goods and services that are better measured and less exposed to international competition. Our VAR analysis produces results that are consistent with these findings, but we do not report them here since they are not necessary to draw our main conclusions.[1]

Together, the first three facts listed above lead us to reject mismeasurement of economic slack, as well as a significant flattening of the wage Phillips curve, as the main cause of the emergence of the inflation–real activity disconnect since 1990. We draw this conclusion because those two explanations are inconsistent with the small change in the dynamic relationship between the most common indicators of cost pressures before and after 1990, at the same time as inflation became much more stable.[2] A further implication of this finding is that we can focus the rest of the investigation on the bivariate relationship between inflation and real activity, without having to take a stance on the most appropriate measure of the latter. Any indicator commonly used in the literature will do.[3]

1. Some recent papers have also explored the behavior of inflation across geographic areas in the United States and across countries (Fitzgerald and Nicolini 2014; Mavroeidis, Plagborg-Møller, and Stock 2014; McLeay and Tenreyro 2019; Hooper, Mishkin, and Sufi 2020; Geerolf 2020). They generally find that the correlation between inflation and unemployment in the cross section is stronger and more stable than in the time series. Hazell and others (2020) provide a guide to translate this cross-sectional evidence into the time series elasticity that is of interest to most of the macroeconomics literature. Using data on US states, they recover a flat Phillips curve once the estimates are properly rescaled, with some evidence of a further reduction in the slope coefficient after 1990. Fully reconciling this evidence across geographies and exchange rate regimes with our conclusion requires more work.

2. We cannot rule out that all the indicators of cost pressures that we include in our analysis have become a poorer proxy for the "true" real marginal costs that drive firms' pricing decisions after 1990. However, it is unlikely that an unobserved change in the dynamics of those costs could have left almost no trace on the joint dynamics of all those indicators.

3. In practice, we focus primarily on the relationship between inflation and unemployment, but we continue to do so in the context of a VAR that also includes other macroeconomic variables. We focus on unemployment because it is arguably the most straightforward and widely discussed measure of the health of the real economy, as well as the most commonly used independent variable in Phillips curve regressions.

This conclusion marks the boundary to which we can push the VAR for purely descriptive purposes. As illustrated in a recent influential paper by McLeay and Tenreyro (2019), the observed relationship between inflation and real activity is the result of the interaction between aggregate demand and supply. The latter captures the *positive* relationship between inflation and real activity, usually associated with the price Phillips curve. Higher inflation is connected with higher marginal costs, which in turn tend to rise in expansions, when demand is high, the labor market is tight, and wages are under pressure. On the contrary, the economy's aggregate demand captures a *negative* relationship between inflation and real activity, which reflects the endogenous response of monetary policy to inflationary pressures. When inflation is high, the central bank tightens monetary policy, thus slowing the real economy. Therefore, the observed cyclical disconnect between inflation and real activity might be the result of either a flat Phillips curve—the slope hypothesis—or a flat aggregate demand, generated by a forceful response of monetary policy to inflation. In the limit in which the central bank pursues perfect inflation stability, inflation is observed to be insensitive to the cycle, regardless of the slope of the Phillips curve. We refer to this second possible explanation for the stability of inflation as the policy hypothesis.

Distinguishing between these two hypotheses is a classic identification problem that requires economic assumptions that were not needed for the data description exercise in the first part of the paper. We impose those restrictions through two complementary approaches, a structural VAR (SVAR) and an estimated dynamic stochastic general equilibrium (DSGE) model. In the SVAR, we identify cyclical fluctuations that can be plausibly attributed to a demand disturbance. To do so, we follow Gilchrist and Zakrajšek (2012) and use their data on the excess bond premium (EBP) to identify a financial shock that propagates through the economy like a typical demand shock, by depressing both real activity and inflation. We choose this shock as a proxy for demand disturbances because it accounts for a significant fraction of the business cycle fluctuations behind the facts described in the first part of the paper. In response to this demand shock, inflation barely reacts in the post-1990 sample, while it used to fall significantly before 1990. This result indicates that the slope of the aggregate supply relationship must have fallen since 1990. Intuitively, the demand shock acts as an instrument for cost pressures in the Phillips curve, identifying its slope. If real activity declines in response to an EBP shock, as it clearly does in both samples, and this lowers cost pressures (i.e., if the instrument is not weak), a muted response of inflation implies a flat Phillips curve.

Although this evidence clearly points in the direction of a very flat aggregate supply curve after 1990, it does not rule out the possibility that monetary policy might have also contributed to the observed stability of inflation. In fact, the main implication of this hypothesis is that monetary policy should lean more heavily against inflation by limiting the impact of demand shocks on the real fluctuations. At the limit in which inflation is perfectly stable, demand shocks should leave no footprint on the real variables. The impulse responses to the EBP shock are far from implying no reaction of the real variables to the demand disturbance, as we would expect if the stability of inflation were due to monetary policy, although they do point to some stabilization, at least in the short run.

The SVAR evidence that we just described helps narrow down the relative contribution of the slope and policy hypotheses for the stability of inflation. To provide an even sharper quantification of their respective roles, we turn to an estimated DSGE model. This exercise is subject to the typical trade-off associated with imposing tighter economic restrictions on the data. On the one hand, we can map the slope and policy hypotheses directly onto parameters of the model that we can estimate on data before and after 1990. On the other hand, the results of this exercise hinge on the entire structure of the model, rather than on a looser set of identifying assumptions as in the SVAR. Therefore, they stand or fall together with the observer's beliefs about the validity of that structure as a representation of the data. To support the case in favor of the model's validity for our purposes, we show that it reproduces the facts generated by the reduced-form VAR used for data description in the first part of the paper. In terms of the two hypotheses, the DSGE estimates point in the direction of a much flatter Phillips curve in the second sample. If we assume that the slope of the Phillips curve is the only parameter that changes after 1990, the estimated model still broadly reproduces the empirical facts. If we only allow policy to change, the estimated model falls short.

Together, the results of the SVAR and DSGE produce two conclusions. First, there is strong support for the slope hypothesis: the slope of the Phillips curve has fallen very substantially after 1990, although it has not gone all the way to zero. Second, there is also some evidence that the policy hypothesis—or other structural changes contributing to a flatter aggregate demand curve—might have contributed to reduce the cyclical sensitivity of inflation, but this evidence is weaker.

The rest of the paper proceeds as follows. The remainder of this section reviews the literature. Section I describes the VAR that we use for data description, the results of which are then described in section II. Section III introduces a stylized aggregate demand and supply framework inspired by McLeay and Tenreyro (2019), which illustrates the fundamental identification problem underlying the interpretation of the observed relationship between inflation and real activity. This model also guides the interpretation of the impulse responses to the EBP shock presented in section IV. Section V revisits the same facts presented in section II from the perspective of an estimated DSGE model and uses that model to further explore the relative contribution of the slope and policy hypothesis to the observed stability of inflation. Section VI elaborates on some policy implications of our main findings, before offering some concluding remarks in section VII.

THE LITERATURE The literature has explored four main classes of explanations for the reduction in the observed correlation between inflation and real activity.

The first set of explanations is related to mismeasurement of either inflation or economic slack. In the inflation dimension, much of the debate has focused on the role of new products and quality adjustment in the construction of price indexes and in the measurement of output and productivity, especially following the introduction of technologies with a very visible impact on everyday life, such as the internet and smart phones (Moulton 2018). This branch of the literature has also explored the recent emergence of online retailing as a source of transformation in firms' pricing practices (Cavallo 2018; Goolsbee and Klenow 2018). By focusing on cyclical fluctuations, our analysis mostly bypasses these considerations, since they primarily pertain to the level of measured inflation. In addition, the inflation–real activity disconnect predates the potential effect of information technology on price mismeasurement, further reducing the potential explanatory power of this hypothesis for our phenomenon of interest.

On the real activity front, the definition and measurement of economic slack have been the subject of a vast literature. Abraham, Haltiwanger, and Rendell (2020) in this volume offer a very recent example. Much of this work has focused on the estimation of potential output and the natural rate of unemployment as reference points to assess the cyclical position of the economy and its influence on inflation. Crump and others (2019) provide a comprehensive discussion of this literature, which features many prominent contributions in the Brookings papers. Our results on the stability of the co-movement of various measures of cost pressures should reduce the weight put on explanations of the inflation disconnect based on the idea that any one measure of slack might be less representative of underlying inflationary pressures after 1990, for instance, due to changes in the relationship between measured unemployment and the overall health of the labor market (Stock 2011; Gordon 2013; Hong and others 2018; Ball and Mazumder 2019).

A second set of explanations for the emergence of the inflation–real activity disconnect focuses on a flatter wage Phillips curve and, more in general, on structural transformations in the labor market and its connection with the goods market (Daly and Hobijn 2014; Stansbury and Summers 2020; Faccini and Melosi 2020). Taken together, our results suggest that whatever structural change might have occurred in the labor market, it is unlikely to be the leading cause of inflation stability. In a recent Brookings paper, Crump and others (2019) capture some of these structural transformations in their Phillips curve estimates by anchoring the inference on the natural rate of unemployment to disaggregated data on labor market flows. This procedure produces a model of inflation that accounts for its dynamics throughout the sample. However, doing so requires a low slope coefficient, as stressed by Davis (2019) and Primiceri (2019) in their discussions.

A third set of explanations focuses on the role of policy in delivering stable inflation. The idea is that a stronger response of monetary policy to inflation flattens the aggregate demand curve, weakening the connection between inflation and real fluctuations, even if the aggregate supply relationship is unchanged (Fitzgerald and Nicolini 2014; Barnichon and Mesters 2019a; Hooper, Mishkin, and Sufi 2020; McLeay and Tenreyro 2019; Kareken and Solow 1963; Goldfeld and Blinder 1972). An implication of this hypothesis is that the Phillips curve is hibernating: a stronger correlation between inflation and business cycles would reemerge if monetary policy reacted less to inflation, as it probably did before the 1990s. Consistent with this view, we also find that monetary policy played some role in stabilizing inflation over the cycle. However, our evidence suggests that policy did not entirely succeed in eliminating demand-driven real fluctuations, implying that it cannot be the dominant driver of the inflation–real activity disconnect. This result, however, leaves open the possibility that changes in monetary (and perhaps fiscal) policy were behind the low-frequency fluctuations in inflation related to its slow rise between the mid-1960s and 1979 and its return to 2 percent over the subsequent two decades, as argued for instance by Primiceri (2006).

Related to this policy hypothesis is the large literature on the role of inflation expectations and their anchoring (Orphanides and Williams 2005; Bernanke 2007; Stock 2011; Blanchard, Cerutti, and Summers 2015; Blanchard 2016; Ball and Mazumder 2019; Carvalho and others 2019; Jorgensen and Lansing 2019; Barnichon and Mesters 2019b). Empirically, expectations are now less volatile than they were before 1990, as we also find in our VAR. However, this observation does not establish that changes in their formation, perhaps in response to shifts in the conduct or communication of policy, represent an autonomous source of inflation stability. Rather, our evidence suggests that the behavior of inflation expectations mostly reflects the inflation stability induced by the flattening of the aggregate supply curve, instead of being its primary source.

In conclusion, our results support a fourth set of explanations that attribute the inflation–real activity disconnect to forces that reduce the response of goods prices to the cost pressures faced by firms, lowering the slope of the structural price Phillips curve. This is the slope hypothesis, which takes several variants. The most prominent is the one that attributes a reduction in the response of prices to marginal costs to the increased relevance of global supply chains, heightened international competition, and other effects of globalization (Sbordone 2007; Auer and Fischer 2010; Peach, Rich, and Linder 2013; Tallman and Zaman 2017; Forbes 2019a, 2019b; Obstfeld 2020; Forbes, Gagnon, and Collins 2020). In a similar vein, Rubbo (2020) points to changes in the network structure of the US production sector.[4]

Compared to this literature that concentrates on estimating the slope of the Phillips curve as a summary statistic of the connection between inflation and real activity, our VAR approach explores more broadly the dynamic relationships among real and nominal variables to draw conclusions on the mechanisms that drive them and how they have changed since 1990. Another advantage of our approach is that it focuses on business cycle dynamics, abstracting from lower-frequency trends and other developments that might be less informative about the Phillips curve relationship. As a result, we do not address the reasons why inflation has been stubbornly below most central banks' targets for the better part of the last decade.

I. Methodology and Data

The objective of this paper is to shed light on the possible causes of the widely acknowledged attenuation of the response of inflation to labor market slack over the past three decades. This section illustrates the methodology and the data that we use to document this fact and its

4. Afrouzi and Yang (2020) connect changes in the conduct of monetary policy directly to the slope of the Phillips curve, straddling the two strands of the literature that we just discussed. They present a model in which rationally inattentive price setters respond less to aggregate shocks when monetary policy is committed to inflation stabilization.

relationship with the behavior of a broad set of macroeconomic variables, whose joint dynamics might help to discriminate among alternative explanations.

I.A. Methodology

To study macroeconomic dynamics and their change over time, we begin by adopting the following vector autoregression (VAR) model:

$$y_t = c + B_1 y_{t-1} + \ldots + B_p y_{t-p} + u_t. \tag{1}$$

In this expression, y_t is an $n \times 1$ vector of macroeconomic variables, which is modeled as a function of its own past values, a constant term, and an $n \times 1$ vector of forecast errors (u_t) with covariance matrix Σ. The reduced-form shocks (u_t) are a linear combination of n orthogonal structural disturbances (ε_t), which we write as $u_t = \Gamma\varepsilon_t$.

VARs are flexible multivariate time series models, which provide a rich account of the complex forms of autocorrelation and cross-correlation that are typical of macroeconomic variables. To synthesize and illustrate these relationships, we study the dynamic response of the variables of interest to a typical unemployment shock. We identify this "U shock" using a Cholesky scheme with unemployment ordered first. This shock corresponds to the linear combination of structural disturbances that drives the one-step-ahead forecast error in unemployment. The impulse responses to this shock tell us how the system evolves in the future if next quarter's unemployment rate turns out to be higher than expected.

The specific combination of shocks responsible for the one-step-ahead forecast error in unemployment accounts for the bulk of business cycle fluctuations in real activity, but it ignores other sources of macroeconomic variation. As a consequence, this part of the analysis has little to say about the substantial share of inflation variability that is independent of the U shocks. Instead, it focuses on the component of inflation that responds to real business cycle impulses, which is the essence of the Phillips curve. Moreover, this approach does not attempt to pin down the precise identity of the structural disturbances driving fluctuations, as in more typical structural VARs. Doing so would require additional economic assumptions, which are not necessary to document many of the empirical facts regarding the attenuated response of inflation to labor market slack that have been discussed in the literature. The advantage of this methodology is that we can illustrate these facts in the context of a unified, dynamic, multivariate statistical framework, without imposing any theoretical restriction.

A limitation of our method is that, without economic restrictions, the facts that it uncovers can be mapped onto more than one hypothesis on the sources of the increased disconnect between inflation and real activity. Therefore, sections IV and V will also explore more economically demanding approaches—SVARs and DSGE models—in order to further pinpoint the source of the empirical observations illustrated below.

The VAR approach to data description that we pursue in this section also has some advantages over the much more popular direct estimation of a Phillips curve, defined as the relationship between inflation (and its lags) and some measure of slack. First, such an inflation equation is embedded in the VAR, which therefore encompasses the single-equation approaches as long as the same variables used in them are included in y_t. Second, embedding such an equation into a multivariate framework explicitly recognizes the challenging identification problem of distinguishing a Phillips curve, which represents the economy's aggregate supply, from its aggregate demand. We illustrate this challenge in the context of a stylized New Keynesian model in section III. Third, looking at the response of inflation and the other variables to the combination of shocks responsible for the one-step-ahead forecast error of unemployment produces more flexible measures of economic slack than those based on specific indicators of potential output or natural unemployment—two notoriously elusive concepts. Fourth, we do not need explicit measures, or a model, of inflation expectations, as long as the variables included in the VAR approximately span the information set used by agents to form those expectations. This aspect of the analysis is especially important, since inflation expectations are a key ingredient in most formulations of the Phillips curve. At the same time, given their relevance, we also consider VAR specifications that include a direct measure of expectations in the vector y_t.

I.B. Data

What variables should the VAR include to provide a comprehensive view of the forces shaping the connection between inflation and the labor market? To answer this question, we refer to an intuitive description of that connection, which is embedded in most formal and informal frameworks built around a price or wage Phillips curve.[5] When firms try to hire more workers to satisfy higher demand for their output, wages tend to rise. Given labor productivity, this increase in wages is associated with higher

5. A simple example of such a framework is the New Keynesian model with sticky wages and prices in chapter 6 of Galí (2015).

marginal costs and inflation. Therefore, a tight labor market and rising wages, costs, and inflation tend to occur together in response to demand shocks.

To characterize these channels, data on inflation and unemployment are not enough. In addition, we need measures of wages, labor productivity, and firms' costs to capture the intermediate steps of the transmission. Therefore, we propose a baseline VAR that includes eight macroeconomic variables: (1) *unemployment*, measured by the civilian unemployment rate; (2) *natural unemployment*, measured by the CBO estimate; (3) *core inflation*, measured by the annualized quarterly growth rate of the personal consumption expenditures (PCE) price index, excluding food and energy; (4) *inflation*, measured by the annualized quarterly growth rate of the GDP deflator; (5) *GDP*, measured by the logarithm of per capita real GDP; (6) *hours*, measured by the logarithm of per capita hours worked in the total economy; (7) *wage inflation*, measured by the annualized quarterly growth rate of the average hourly earnings of production and nonsupervisory employees (PNSE); and (8) the *labor share*, measured by the logarithm of the share of labor compensation in GDP.

Besides unemployment, this VAR includes a block of variables referring to the total economy: GDP, hours, the labor share, and the GDP deflator. These variables can be combined to compute a measure of hourly nominal compensation in the total economy. The growth rate of this measure of nominal wages closely tracks compensation per hour in the nonfarm business sector, a commonly used indicator of labor costs. The problem with both these series is that they are extremely volatile at high frequencies, obscuring the underlying wage dynamics over the business cycle. For this reason, the baseline VAR also includes the PNSE wage inflation series. This series only covers about 80 percent of private industries, but it is substantially less noisy than the more comprehensive ones that we mentioned. Finally, in addition to GDP inflation, our model also includes core PCE inflation, given its importance as a gauge of underlying inflationary pressures.

We estimate this eight-variable VAR over two nonoverlapping samples, to investigate possible changes in the typical co-movement pattern of these variables in response to the U shock described in section I.A. The first sample ranges from 1964:Q2 to 1989:Q4, and the second from 1989:Q1 to 2019:Q3.[6] The analysis starts in 1964:Q2, when the PNSE wage inflation

6. The first four observations are used as initial conditions, since our VAR has four lags. Therefore, effectively, the estimation starts in 1965:Q2 in the first sample and in 1990:Q1 in the second one.

series first becomes available. The date at which we split the sample is the result of a compromise. On the one hand, there is some evidence that macroeconomic dynamics might have changed around 1984, after the first phase of the so-called Volcker disinflation. On the other hand, a simple inspection of the data suggests that inflation has been most stable starting around the mid-1990s, after the opportunistic disinflation engineered by the Federal Reserve under Alan Greenspan following the 1990–1991 recession. As a compromise between these two alternatives, we split the sample in 1990. This choice also has the advantage of creating two samples of fairly similar lengths. Section A of the online appendix shows that this choice is immaterial for the results.

The data are quarterly and the VAR includes four lags. It is estimated with Bayesian methods and a standard Minnesota prior, given the relatively high number of variables and short sample sizes. The tightness of the prior is chosen based on the data-driven method described in Giannone, Lenza, and Primiceri (2015).

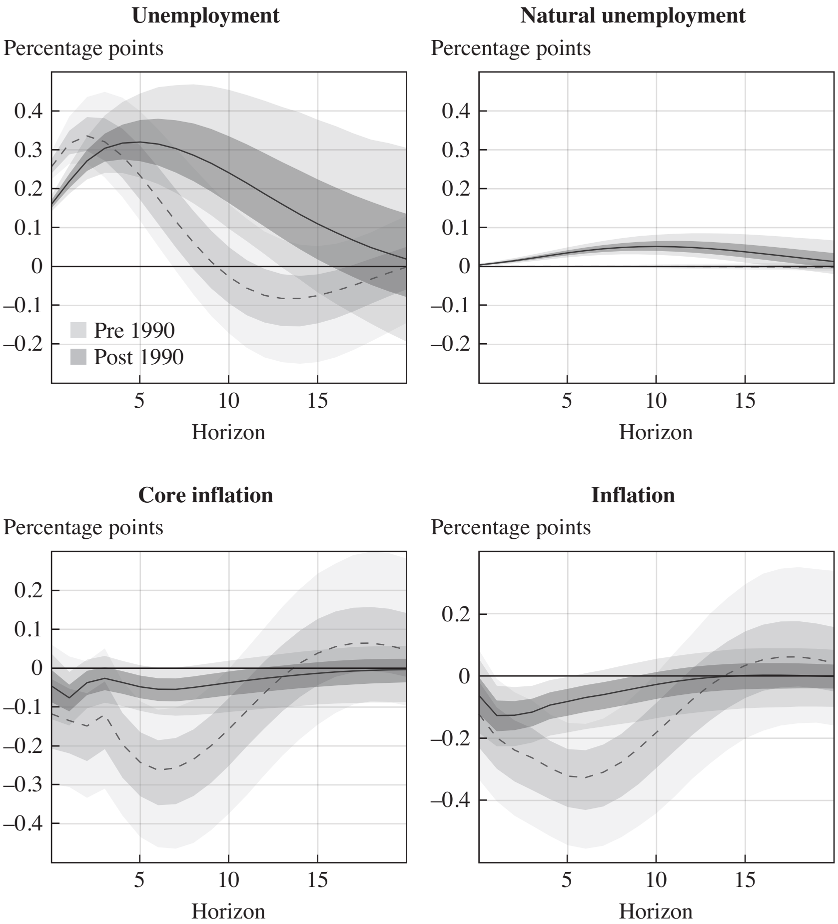
II. Facts

This section documents a number of known and new facts concerning the relationship between unemployment, inflation, and some other key macroeconomic variables. These empirical findings lead us to two main conclusions: the attenuation of inflation fluctuations over the business cycle before and after 1990 is striking, and, in comparison, the co-movement of all real variables and indicators of firms' cost pressures has been remarkably stable.

II.A. Fact 1: Unemployment and Price Inflation

We begin by presenting the impulse responses of unemployment and inflation to a U shock in the two samples. Figure 1 shows that inflation falls significantly as unemployment rises in the first sample. This finding suggests that the U shock is characterized by a strong demand component, which explains why traditional Phillips curves estimated over this sample have a negative slope. In the second sample, instead, unemployment increases by a roughly similar amount, but the responses of both inflation measures are muted. In fact, the response of core inflation is statistically indistinguishable from zero throughout the horizon, while that of GDP inflation is a bit more negative and borderline significant after about one year. In addition, the very flat response of natural unemployment

Figure 1. Impulse Responses to an Unemployment Shock in the Baseline VAR



Source: Authors' calculations.

Note: The impulse responses are from the baseline VAR described in section I.B. The shock is identified using a Cholesky strategy, with unemployment ordered first. The solid lines are posterior medians, while the shaded areas correspond to 68 percent and 95 percent posterior credible regions. The pre- and post-1990 samples consist of data from 1964:Q2 to 1989:Q4 and from 1989:Q1 to 2019:Q3, respectively.

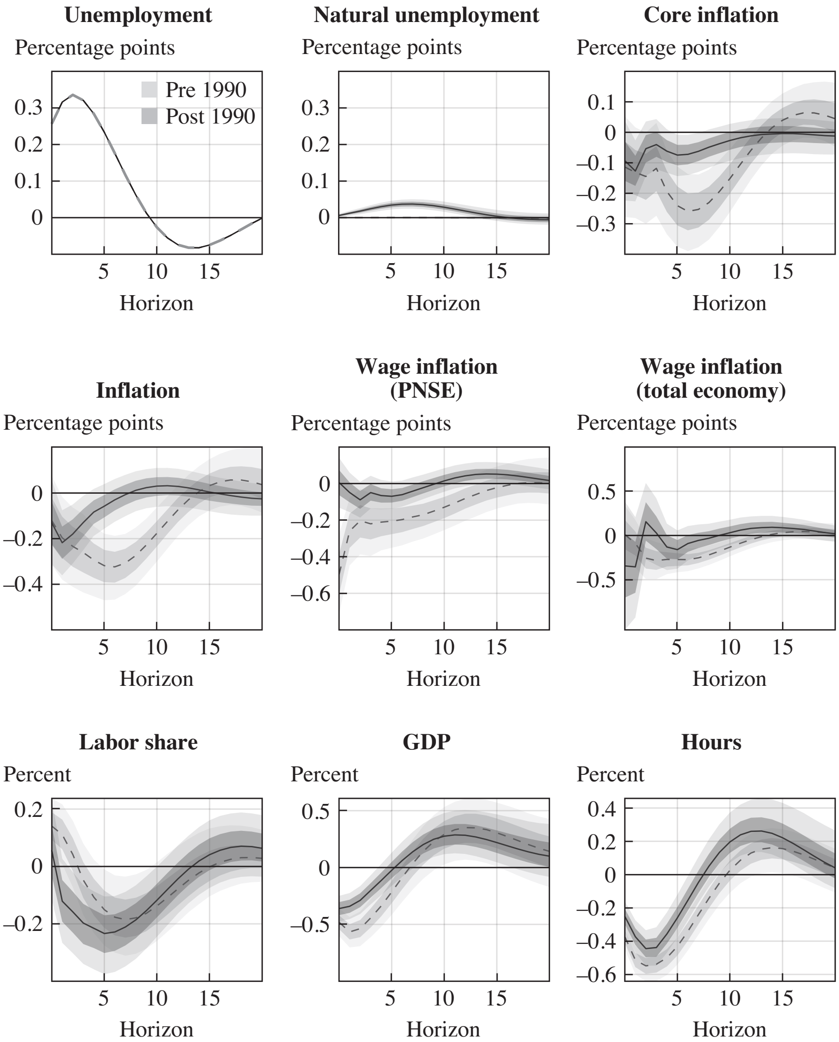
indicates that the shock only captures business cycle variation. Therefore, looking at the reaction of unemployment or the unemployment gap to this shock would produce identical results.

Online appendix A shows that the responses of figure 1 are nearly identical to those to a typical business cycle shock, obtained as the linear combination of structural disturbances that drives the largest share of unemployment variation at business cycle frequencies, as in Giannone, Lenza, and Reichlin (2019) and Angeletos, Collard, and Dellas (2019). This result suggests that the combination of shocks associated with the one-step-ahead forecast error in unemployment and the one responsible for the bulk of business cycle fluctuations are virtually the same. Our finding also casts some doubt on the interpretation of the muted response of inflation to business cycle shocks proposed by Angeletos, Collard, and Dellas (2019), since they do not explain why such response was much more vigorous before 1990.

The response of unemployment to a U shock in figure 1 is more persistent in the second sample. This feature of economic fluctuations is evident even from the raw data, and it is consistent with the lengthening of expansions in the last thirty years. However, this change in the profile of unemployment fluctuations does not play much of a role in accounting for the attenuated response of inflation in the second sample. We illustrate this point with an exercise that forces the response of unemployment to be identical in the two samples. Specifically, we compute the responses of all variables as the difference between their forecast conditional on a specific path of unemployment and their unconditional forecast, following the methodology of Bańbura, Giannone, and Lenza (2015). As this common path in both samples, we choose the median response of unemployment to a U shock in the first sample.[7] Figure 2 plots the dynamics of all the VAR variables in this conditional forecast exercise. As in figure 1, the response of inflation in the second sample is much attenuated, although it now remains negative. Online appendix A shows that this change in inflation dynamics over the two samples is not limited to the two measures of inflation included in the baseline VAR, but it extends to a number of other commonly used inflation series.

7. This conditional forecast approach recovers the most likely sequence of shocks to guarantee that unemployment follows a given path. In this respect, it has a slightly different interpretation relative to the impulse responses, because the latter are based on a single shock perturbing the economy at horizon zero.

Figure 2. Impulse Responses Conditional on a Path for Unemployment in the Baseline VAR



Source: Authors' calculations.

Note: These responses are computed by applying the methodology described in section II.A to the baseline VAR of section I.B. The solid lines are posterior medians, while the shaded areas correspond to 68 percent and 95 percent posterior credible regions. The pre- and post-1990 samples consist of data from 1964:Q2 to 1989:Q4 and from 1989:Q1 to 2019:Q3, respectively.

These findings can be summarized into a first key stylized fact: the sensitivity of goods price inflation to labor market slack has decreased dramatically after 1990. This fact provides a complementary, more dynamic, characterization of many findings in the literature regarding the stability of inflation. Interpreting this fact is the main task of the rest of the paper.

II.B. Fact 2: Unemployment and Wage Inflation

A substantial body of recent work finds that the connection of wage inflation to labor market slack remains stronger than that of goods inflation (Galí and Gambetti 2019; Hooper, Mishkin, and Sufi 2020; Rognlie 2019). This section presents VAR results broadly consistent with these findings. The second row of figure 2 plots the response of two measures of nominal wage inflation using the conditional forecast approach described in the previous subsection. The first measure (PNSE, middle graph) is the one used directly for the estimation of the baseline VAR. The second (total economy, right graph) is that implied by the data on the labor share, hours, output, and GDP inflation.[8] The reaction of the PNSE series is attenuated in the post-1990 period, while the response of the total economy measure shows more similarities in the two samples. Therefore, we take the balance of the evidence as consistent with the view that the connection between wage inflation and unemployment remains alive, although it is weaker in the more recent period. As shown in online appendix A the sensitivity of wage inflation to unemployment after 1990 is even stronger when wages are measured with the employment cost index (ECI), which is arguably a better measure of the cyclicality of wages than the ones used in this section. Unfortunately, the ECI is only available starting in 1980, preventing a full comparison of its behavior pre- and post-1990. We summarize these findings in the form of a second stylized fact: the sensitivity of nominal wage inflation to labor market slack has diminished after 1990, but less than that of price inflation.

One implication of this fact is that explanations of the unemployment-inflation disconnect involving a much reduced responsiveness of wage inflation to labor market slack are not very plausible. For example, a popular narrative attributes the stability of inflation during the Great Recession to the existence of downward nominal wage rigidities: if firms are reluctant or unable to lower nominal wages, their marginal costs should remain

8. In logs, the labor share (ls) is defined as the sum of the nominal wage (w) and hours (h), minus real GDP (gdp) and the price level (p), or $ls \equiv w + h - gdp - p$. Therefore, this measure of the (log) nominal wage is constructed as $w = ls - h + gdp + p$.

relatively high, putting upward pressure on prices and inflation. Such a story, however, would imply a substantial weakening of the co-movement between unemployment and wage inflation, which seems at odds with the data. In addition, as we demonstrate in online appendix A, this co-movement is approximately equally strong after 1990 regardless of whether we include or exclude the Great Recession period.

II.C. Fact 3: Unemployment and Unit Labor Costs

One obvious difficulty in interpreting the evidence on the connection between nominal wage inflation and unemployment presented in the previous section is that it also partly reflects a weaker response of goods inflation. Mechanically, nominal wage inflation is the sum of real wage inflation and goods inflation. Therefore, the former will appear less responsive to the cycle if the latter reacts less given the dynamics of the real wage. A more helpful approach to evaluate the implications of wage dynamics for inflation, therefore, is to study more direct measures of how wages contribute to firms' marginal costs.[9] The most popular proxy for aggregate real marginal costs are unit labor costs (or, equivalently, the labor share). With constant returns in production, (log) unit labor costs are proportional to (log) marginal costs. Under more general assumptions, this proportionality no longer holds, but unit labor costs are likely to remain a more accurate gauge of the cost pressures faced by firms than nominal wage inflation.[10]

Figure 2 shows that the forecast of the labor share conditional on the usual path of unemployment is very similar in the two samples. This observation leads to the third stylized fact: the co-movement of unemployment and the labor share over the business cycle is stable over time. This fact supports and further refines the view according to which labor market developments are unlikely to be the main source of the change in inflation dynamics over the past thirty years. The claim is not that labor market

9. In pricing problems based on cost minimization, firms' marginal costs are the key driver of their pricing decisions. As a result, the evolution of aggregate marginal costs is the fundamental source of inflation in a large class of models with nominal rigidities. These models include those with staggered price setting, as in Calvo (1983) and Taylor (1980), as well as those with sticky information or rational inattention, as in Mankiw and Reis (2002) or Mackowiak and Wiederholt (2009).

10. With constant returns to scale production, a firm's log marginal cost is proportional to its log unit labor cost, defined as $ulc = w - (gdp - h)$. With homogeneous factor markets, marginal cost is equalized across firms, so that the *aggregate* log labor share ($ls = w + h - gdp - p$) is proportional to the average real marginal cost.

dynamics have not changed since 1990. More narrowly, the statement is that, whatever those changes might have been, they did not have a significant impact on the dynamics of firms' marginal costs, at least as seen through the lens of a proxy such as the labor share. The next section adds one further dimension to this claim by showing that the same can be said of other well-known aggregate proxies for firms' cost pressures.

II.D. Fact 4: Unemployment and Other Measures of Real Activity

The previous subsection argued that unit labor costs are likely to be the most informative variable on the extent to which cost pressures originating in the labor market are transmitted to goods prices.[11] Next, we show that the dynamics of many other variables used in the literature to capture real sources of inflationary pressure, from the labor market or otherwise, are also relatively stable over time. The third row of figure 2 reports the conditional forecasts of hours and output. These responses are essentially identical over time, implying a fourth stylized fact: the business cycle correlation among several indicators of real activity has not changed in the two samples. Online appendix A further shows that these empirical patterns also hold for the output gap and the employment-to-population ratio, when we add these variables to the baseline VAR.

The important conclusion that we draw from these results is that the severe illness of the reduced-form relationship between inflation and real activity cannot be cured by picking a different indicator of either labor or goods market slack among those commonly used in the literature. In fact, the remarkable stability in the dynamic relationships between all the real variables that we have considered suggests that the diagnosis of what ails inflation should be independent of one's view on the best proxy for underlying inflationary pressures.

II.E. Adding Interest Rates and Expected Inflation

In this subsection, we augment the model with data on the federal funds rate and on long-term inflation expectations from the survey of professional

11. A key implication of firms' cost minimization is that marginal cost is equalized across all inputs. As a result, marginal cost pressures—measured by comparing wages to labor productivity—provide a comprehensive view of the cost pressures faced by firms, even if the input whose direct cost is rising is not labor. The main difficulty in operationalizing this observation is the measurement of the marginal cost and marginal benefit of labor, i.e., the wage and the marginal product of labor. Available measures of wages and (average) labor productivity capture those marginal concepts only under restrictive assumptions.

forecasters (SPF).[12] The former was not included in the baseline VAR because it was at the zero lower bound (ZLB) for many years in the second sample. To avoid that period, this larger VAR is estimated excluding data after 2007. Figure 3 plots the conditional forecast of all the variables in the model. Compared to the baseline, the conditioning path of unemployment, which is, as usual, the median response of unemployment to the U shock in the pre-1990 sample, returns to zero faster, although its inverted *S* shape is otherwise similar. Moreover, the estimated responses are more uncertain in the second sample, since it is shorter by about twelve years. However, the main empirical facts documented so far are robust to these changes.

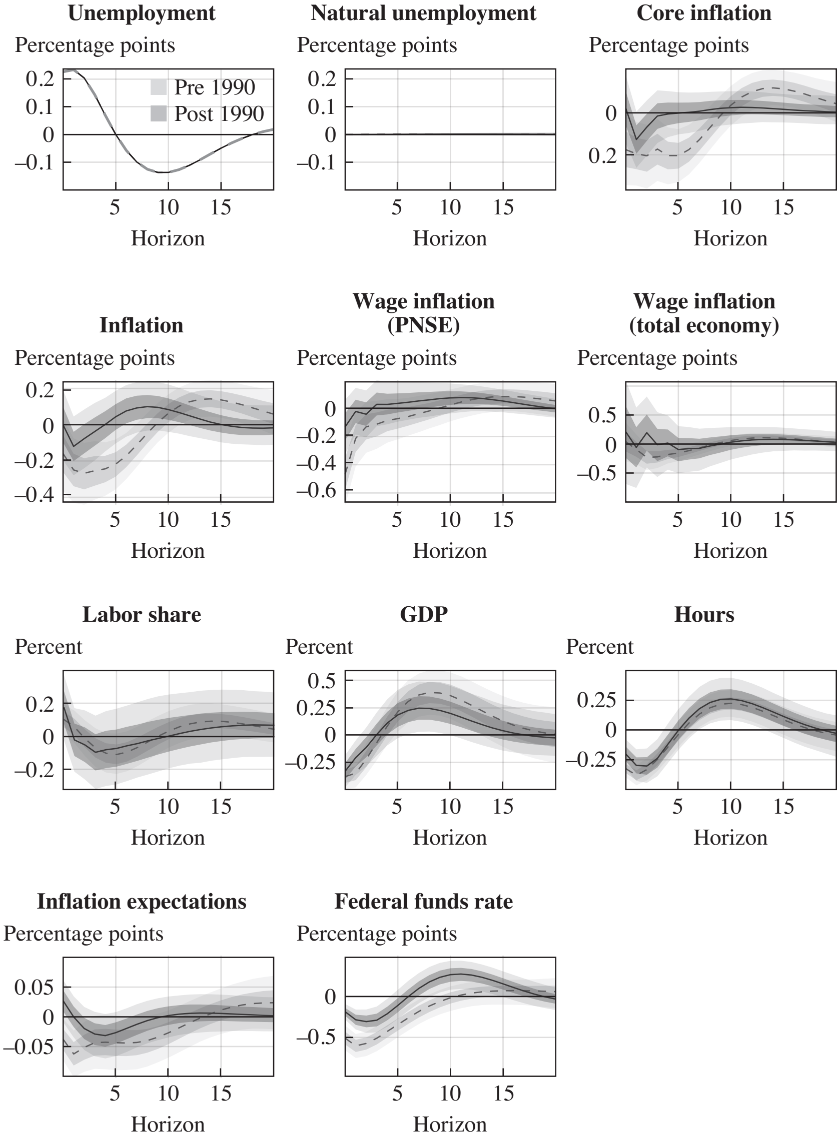
Focusing on the newly added variables, the response of the federal funds rate in the two samples has a similar shape, but it is less persistent after 1990. That of inflation expectations is more muted in the second sample, similar to inflation. At the same time, the gap between the two variables falls significantly in the first sample, while it is more stable in the second, just as is the case with inflation itself. This observation suggests that the reduction in the sensitivity of inflation to business cycles goes beyond what can be explained through the increased stability of long-run inflation expectations. The extent to which more-anchored expectations simply reflect the increased stability of inflation, as opposed to being one of its independent sources, remains an open question. We will return to this issue in section IV.

II.F. Summary of the Key Facts

The four stylized facts documented above lead us to two important conclusions, which crucially inform the rest of the analysis. First, the change in the business cycle dynamics of inflation before and after 1990 stands out compared to that of all the real variables that we have considered. Second, the co-movement of these real variables is remarkably similar before and after 1990. Together, these two observations suggest that we can focus the rest of the analysis on the bivariate relationship between inflation and real activity, with no need to be more specific on its measurement. As illustrated in section III, however, this significant narrowing of the scope of the inquiry is not sufficient to conclude that the anemic response of inflation to the cycle is due to a flattening of the structural Phillips curve. The reason is that a flattening of the aggregate demand relationship, perhaps

12. The data on long-term inflation expectations are constructed as in Clark and Doh (2014) and Del Negro and others (2017).

Figure 3. Impulse Responses Conditional on a Path for Unemployment in a Larger VAR



Source: Authors' calculations.

Note: These responses are computed by applying the methodology described in section II.A to the baseline VAR of section I.B, augmented with long-term inflation expectations and the federal funds rate. The solid lines are posterior medians, while the shaded areas correspond to 68 percent and 95 percent posterior credible regions. The pre- and post-1990 samples consist of data from 1964:Q2 to 1989:Q4 and from 1989:Q1 to 2019:Q3, respectively.

induced by a more forceful reaction of monetary policy to inflation, could in principle result in more stable inflation. Further distinguishing between these two possibilities requires putting more structure on the problem, as we will then do in sections IV and V.

III. Lessons from a Stylized Model

To aid in the interpretation of the empirical facts described in section II, we now introduce a stylized model of the joint determination of inflation and real activity. This model, which is directly inspired by McLeay and Tenreyro (2019), is based on the textbook New Keynesian framework of Woodford (2003) and Galí (2015). However, its implications for the nature of business cycles under alternative hypotheses regarding the possible sources of inflation stability are quite general, as we argue below. We use this simple model to make three essential points: (1) the empirical facts of section II are consistent with two possible explanations of the stability of inflation after 1990: either a reduction in the sensitivity of pricing decisions to marginal cost pressures or a change in the conduct of monetary policy; (2) the key implication that differs across these two hypotheses is that real activity is driven predominantly by demand-type shocks in the first case and supply-type (or cost-push) disturbances in the second; and (3) unfortunately, it is difficult to empirically verify which shocks—demand or supply—are prevalent in the post-1990 period based on the co-movement pattern between inflation and real activity, because the variation of inflation is minimal. Therefore, next we will introduce more information and structure to further sharpen our inference.

III.A. A Simple Model of Aggregate Demand and Supply

The stylized model we consider consists of the following three familiar equations:

$$\pi_t = \beta E_t \pi_{t+1} + \kappa (x_t + s_t), \tag{2}$$

$$x_t = E_t x_{t+1} - \sigma (i_t - E_t \pi_{t+1} - \delta_t), \text{ and} \tag{3}$$

$$i_t = E_t \pi_{t+1} + \psi_\delta \delta_t + \psi_\pi \pi_t, \tag{4}$$

where π_t represents price inflation, $x_t \equiv y_t - y_t^n$ is the output gap (defined as the log deviation of output from a measure of potential), i_t is the nominal interest rate, and s_t and δ_t are exogenous disturbances. In this formulation, $E_t\pi_{t+1}$ and E_tx_{t+1} denote rational expectations of next period's inflation and output gap.[13]

Equation (2) is the model's structural Phillips curve, an aggregate supply relationship that maps higher output gaps into higher inflation. It is based on the dependence of inflation on firms' marginal costs—a fairly general feature of optimal pricing problems (Sbordone 2002)—in combination with some simplifying assumptions that make marginal costs proportional to the output gap. These simplifying assumptions, however, are not restrictive for our analysis, given that facts 3 and 4 in section II document a stable dynamic relationship between many real activity variables usually employed to measure slack, such as unemployment, hours, GDP, and unit labor costs. Therefore, given this stability, we do not need to take a strong stance on what x_t precisely represents in our stylized model and will simply refer to it generically as real activity, or the output gap. Finally, the supply, or cost-push shock (s_t) in equation (2) stems from fluctuations in desired markups, which explains why it is scaled by the slope κ.[14]

The other two equations constitute the demand block of the model. In particular, equation (3) is an Euler equation, or dynamic investment-savings (IS) equation, which connects the nominal interest rate to real activity. The strength of this negative relationship is governed by the parameter $\sigma > 0$. In addition, the equation is perturbed by the shock δ_t, which can be interpreted as capturing fluctuations in the Wicksellian natural rate of interest, due to technology or demand disturbances. We will refer to it as a demand shock, for short. Equation (4) is a simple interest rate rule that represents the response of the monetary policy authority to economic developments. This specification allows for a direct response of the policy rate to the IS shock—for reasons that will become clear shortly—and to inflation, where the Taylor principle requires $\psi_\pi > 0$. Adding a term in the

13. Lowercase letters denote logs, so that, for instance, $y_t \equiv \log Y_t$, where Y_t is the level of output. Y_t^n is natural output, the level of output that would be observed in the absence of nominal rigidities.

14. Under Calvo pricing, fluctuations in desired markups have the same effect on inflation as those in real marginal costs. Therefore, at the limit in which prices never change, the sensitivity of inflation to both real activity and desired markups captured by the parameter κ goes to zero and inflation becomes perfectly stable. More generally, we could allow for other sources of exogenous supply shocks, which might include, for instance, exogenous shifts in inflation expectations that are not fully captured by the rational expectations term $E_t\pi_{t+1}$. In the presence of such shocks, the variability of inflation is not zero even when $\kappa = 0$, but the qualitative implications of the model described below do not change.

output gap, as in Taylor (1993), or a monetary policy shock would not change the model's key qualitative implications.

Plugging the policy rule into the IS equation produces a negative relationship between inflation and real activity of the form

$$\pi_t = -\phi(x_t - E_t x_{t+1} - d_t),$$

where $\phi \equiv (\sigma\psi_\pi)^{-1} \geq 0$ and $d_t \equiv \sigma(\psi_\delta - 1)\delta_t$ is a simple rescaling of the demand shock. If the demand disturbance is observable, monetary policy can perfectly offset it by setting $\psi_\delta = 1$. More generally, demand shocks are likely to be transmitted to the economy at least partially, either because the monetary authority observes them with noise, or because it chooses not to react to them fully, or because it is prevented from doing so by the ZLB. All of these scenarios are captured in this model by $\psi_\delta < 1$, which implies some pass-through of these shocks into inflation.

The negative slope of this aggregate demand equation reflects the fact that monetary policy leans against inflation by raising the real interest rate, which in turns lowers real activity. This feature of aggregate demand does not depend on the exact specification of the interest rate rule, as long as the real interest rate responds positively to inflation. As shown by McLeay and Tenreyro (2019), this is also a feature of aggregate demand under optimal monetary policy. In this respect, our approach to modeling monetary policy and that of McLeay and Tenreyro (2019) are isomorphic, even if they derive the aggregate demand equation directly from an optimal policy problem, without relying on the IS equation. In comparison, our setup with an IS equation and a policy rule is more explicit about some potential sources of demand shocks, but its key implications are the same.

In sum, at a high level of generality, the model is just an aggregate supply (AS) and aggregate demand (AD) framework, similar to those typically found in intermediate macroeconomics textbooks, such as Jones (2018). In fact, most of the intuition that we derive from this framework stems exactly from this underlying demand and supply structure, as in McLeay and Tenreyro (2019).

III.B. Two Alternative Sources of Inflation Stabilization

Given the structure of the model, it is immediately apparent that stable inflation can be the result of at least two changes in the economy: first, a flat structural Phillips curve, which corresponds to $\kappa \to 0$; second, a very elastic aggregate demand curve, which corresponds to $\phi \to 0$ or, in terms of the interest rate rule, to $\psi_\pi \to \infty$. In what follows, we will take these two extreme parametric restrictions as stylized representations of the two alternative hypotheses on the ultimate source of the observed inflation stability that have been most discussed in the literature.[15]

The first hypotheses—the (Phillips curve) slope hypothesis, for short—is that inflation is stable because changes in the structure of goods markets or in firms' pricing practices have produced a structural disconnect between inflation and marginal cost pressures. The literature has explored many mechanisms that might lead to such a disconnect, as reviewed in the introduction. Distinguishing among them is beyond the scope of this paper.

The second hypothesis that we focus on—the policy hypothesis, for short—is that inflation is stable because monetary policy now leans more heavily against inflation than it did in the first part of the sample, thus reducing its variability in equilibrium. This is the hypothesis favored by McLeay and Tenreyro (2019). In our stylized model, this hypothesis amounts to assuming that ψ_π has increased in the second part of the sample. In the limit with $\psi_\pi \to \infty$, inflation becomes perfectly stable. In practice, there are many channels through which a change in the actual conduct of monetary policy and in the communication and public understanding of its objectives can affect inflation dynamics and inflation expectations, without going to the extreme of promising a very large increase in policy rates in reaction to even small changes in inflation.[16]

15. In our stylized model, a flat aggregate demand curve could also result from $\sigma \to \infty$, although its intercept (and thus inflation) would still depend on the demand shock in this case. We do not focus on this possibility because there is not much evidence that the responsiveness of the real economy to interest rates has increased since 1990. If anything, there is some discussion of a reduced pass-through of interest rates, and of financial conditions more generally, onto real variables, especially since the financial crisis.

Another obvious possibility is that the volatilities of both shocks have fallen dramatically. Although the large body of literature on the Great Moderation indeed suggests that the volatility of (at least some) shocks did fall in the mid-1990s, we do not focus on this possible explanation because the volatility of real variables has not fallen nearly as much as that of inflation, at least in response to business cycle shocks. In this respect, our main object of inquiry is the reduction in the volatility of inflation relative to that of its plausible real drivers, conditional on business cycle shocks.

16. Inflation expectations do not play a crucial independent role in our model because under rational expectations they are a function of the same shocks that drive inflation. In this respect, stable inflation and inflation expectations are two manifestations of the same phenomenon. Carvalho and others (2019) discuss the notion of expectation anchoring theoretically and empirically in the context of a model with learning, in which expectations are not as tightly linked to actual inflation as under rational expectations. Jorgensen and Lansing (2019) also study the implications of a learning model for the observed connection between inflation and real activity.

The rest of this section derives some basic implications of these two alternative hypotheses in the context of our stylized model, and discusses the extent to which they are consistent with the evidence in section II.

III.C. Model Solution

This section presents the solution of the simple model described above under the assumption of independent and identically distributed shocks. With this simplifying assumption, expectations are zero and the model reduces to a static demand and supply framework with stochastic shocks,

$$\pi_t = \kappa(x_t + s_t)$$

$$\pi_t = -\phi(x_t + d_t),$$

whose solution is

$$x_t = \frac{\phi}{\phi+\kappa} d_t - \frac{\kappa}{\phi+\kappa} s_t$$

$$\pi_t = \frac{\phi\kappa}{\phi+\kappa} d_t + \frac{\phi\kappa}{\phi+\kappa} s_t.$$

The particular form of this solution is, of course, model specific, but its economics are simple and quite general. Demand shocks induce a positive correlation between inflation and the output gap. With direct observations on d_t, or an instrument for it, it would be possible to estimate the slope of the Phillips curve κ by comparing the response of x_t and π_t to the shock d_t, as in Barnichon and Mesters (2019a, 2019b). On the contrary, supply shocks induce a negative correlation between inflation and the output gap, from whose strength we could infer the demand parameter ϕ. When demand and supply shocks cannot be directly observed, the correlation between inflation and the output gap is not informative on either ϕ or κ, as in the classic identification problem. This is the basic point nicely illustrated by McLeay and Tenreyro (2019).

To further clarify this identification challenge, and to shed light on how to potentially overcome it, consider the solution of the model under the two alternative sources of inflation stabilization that we discussed above, $\kappa \to 0$ or $\phi \to 0$. Inflation is zero in both cases, but $x_t = d_t$ under the slope hypothesis, while $x_t = -s_t$ under the policy hypothesis. In other words, when the slope of the structural Phillips curve is zero—for example, because prices are insensitive to marginal cost pressures—the economy becomes more Keynesian, and demand shocks are the predominant drivers of output fluctuations. On the contrary, when policy leans very heavily against inflation, the economy tracks the flexible price equilibrium, it becomes more neoclassical, and economic fluctuations are driven by supply or cost-push disturbances. In sum, which hypothesis—slope or policy—is a better explanation of post-1990 inflation stability simply depends on whether post-1990 business cycles were mainly driven by supply or demand disturbances.

A popular approach to distinguish between demand- and supply-driven fluctuations is to exploit the co-movement pattern of real activity and inflation. A strongly positive correlation would signal the prevalence of demand shocks, while a negative one would favor the predominance of supply innovations. Unfortunately, this strategy is not effective in our case, given the observed stability of inflation: if inflation varies very little over the business cycle—as it has since the 1990s—it also carries limited information to help us separate demand from supply shocks based on the sign of its co-movement with real activity. This is why, in the next sections, we will attempt to tackle this identification challenge by bringing either more information to the table or more theoretical restrictions on the impact matrix, of the form provided, for instance, by full-blown DSGE models.

IV. Interpreting the Facts with a Structural VAR

The U shock employed in section II is a useful descriptive tool, which helps to focus the empirical analysis on the dynamics of inflation and real activity occurring over the cycle. This exercise focuses on the frequencies at which the connection between the nominal and real side of the economy is usually thought to be most evident, as well as those at which monetary policy might have the most significant impact on these dynamics. But most business cycle models with multiple shocks suggest that these dynamics reflect the responses of the economy to a mixture of structural shocks, even if one of them might be preponderant (Smets and Wouters 2007; Justiniano, Primiceri, and Tambalotti 2010). In terms of the stylized model presented above, the U shock would be a combination of demand and supply disturbances, with weights that depend on the relative variance of those shocks, as well as the structural parameters of the economy, including the slope of the aggregate demand and supply equations. As argued in section III, it is therefore impossible to determine the main source of inflation stability—

a flat aggregate supply or demand—unless we can distinguish the two kinds of shocks more precisely. This is the task of this section.

More specifically, we use data on the excess bond premium (EBP) constructed by Gilchrist and Zakrajšek (2012) to identify a credit market disturbance, which we interpret as a proxy for demand shocks. To do so, we add the EBP to the baseline VAR of section I and study the impulse responses to innovations to the EBP that are orthogonal to the other variables in the system.[17] The idea is that innovations to the EBP capture disruptions in credit markets that propagate through the rest of the economy largely as demand shocks. When credit is tight, as signaled by a high EBP, investment falls, reducing aggregate demand and generating further reactions in the economy that also lead to lower labor demand, lower wages, lower income, and ultimately lower inflation.[18]

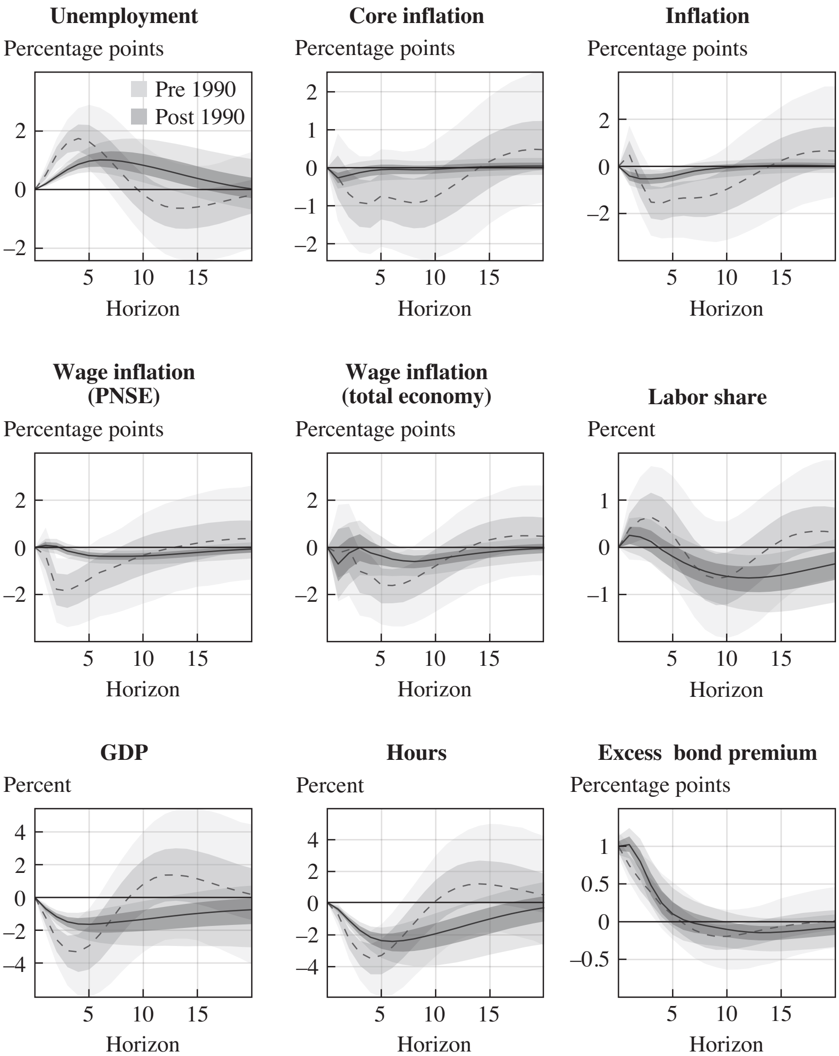
This strategy does not hinge on the identification of genuinely exogenous credit supply shocks, which can be hard to disentangle from other disturbances affecting financial markets, such as uncertainty shocks or even monetary policy shocks (Caldara and others 2016). All we need is that these innovations to the EBP propagate through the economy by shifting primarily the demand for labor and goods, regardless of their ultimate origin, and that this is true to roughly the same extent before and after 1990. This is the maintained assumption in the rest of this analysis.

Figure 4 presents the impulse responses to the EBP innovation described above (to save space, we omit the response of natural unemployment, given that it is flat). The EBP shocks are more volatile in the second sample, mostly reflecting the sharp spike in credit spreads during the financial crisis. Their standard deviation is approximately 60 percent higher after 1990 than before. Therefore, we normalize the size of the shock in both samples so that it increases the EBP by 1 percentage point on impact at the

17. In practice, we order the EBP last in the VAR and use a recursive scheme to identify the shock to the EBP equation. The results are similar if we order the EBP first in the Cholesky ordering, or if we compute impulse responses to the combination of shocks with the highest contribution to the EBP's variance at business cycle frequencies. This identification strategy is similar to that pursued by Gilchrist, Yankov, and Zakrajšek (2009) and Gilchrist and Zakrajšek (2012). Their dynamic systems, a factor-augmented vector autoregression (FAVAR) and a VAR, respectively, also include "fast moving" variables, such as asset prices, which they place below the EBP in the Cholesky ordering. We do not have such variables in our system, so we order the EBP last.

18. This is how marginal efficiency of investment shocks in Justiniano, Primiceri, and Tambalotti (2011), risk shocks in Christiano, Motto, and Rostagno (2014), and spread shocks in Cai and others (2019) propagate. All these shocks are identified mostly through their effect on credit spreads.

Figure 4. Impulse Responses to an EBP Shock



Source: Authors' calculations.

Note: The EBP shock is identified by assuming that it affects the excess bond premium contemporaneously, but all other variables with a lag. The impulse responses are from the baseline VAR of section I.B, augmented with data on the EBP. The solid lines are posterior medians, while the shaded areas correspond to 68 percent and 95 percent posterior credible regions. The pre- and post-1990 samples consist of data from 1973:Q1 to 1989:Q4 and from 1989:Q1 to 2019:Q3, respectively.

median draw. After this normalization, the response of the EBP to its innovation has the same shape in the two samples, which simplifies the evaluation of the changes in the reaction of the other variables.

Among these variables, the inflation rates barely react in the second sample, consistent with the findings in section II. According to the intuition developed within the stylized model presented above, most of the information to distinguish between the slope and policy hypotheses should come from the responses of the real variables. Under the policy hypothesis, the economy should become more neoclassical, with demand shocks having smaller effects on the real variables. On the contrary, under the slope hypothesis the economy should become more Keynesian, with demand shocks becoming more destabilizing. Comparing the second sample to the first, there is evidence of an attenuated response of unemployment, GDP, and hours in the first few quarters after the shock. This piece of evidence is consistent with the policy hypothesis, especially considering that this is the horizon at which monetary policy has arguably the most bite on the real economy. However, the response of all the real variables is much more persistent in the second sample. For instance, the post-1990 response of unemployment remains statistically and economically positive for a substantially longer period of time. As a consequence, the effect of the EBP shock on unemployment, cumulated over a five-year horizon, is actually overall larger in the second sample than in the first. Similar considerations hold for GDP, hours, and the labor share.

On balance, this exercise provides fairly strong evidence in favor of the slope hypothesis, given that the response of inflation has become much more muted than that of the real variables. However, the experiment does not completely rule out an important contribution of monetary policy in better insulating the economy from demand shocks, and hence delivering more stable inflation. Parsing this evidence into sharper conclusions on the relative contribution of these two developments to the stability of inflation since 1990 is very difficult to do without putting more structure on the identification problem. This is what we do in the next section.

An alternative, perhaps more intuitive, way of presenting these results is to retell them through the perspective of the Great Recession. Most observers agree that the Great Recession originated from the financial crisis that preceded it. Although that shock had a complex origin and it affected the economy through many channels, it had the hallmarks of a typical demand shock. In our VAR, one of the main manifestations of this shock is a massive increase in the EBP. If this shock was indeed primarily a demand shock, the fact that inflation fell by a very limited amount is a strong indication that the slope of the Phillips curve must be very low, at least relative to what it used to be before the 1990s.

At the same time, the fact that the economy weathered the storm without a collapse more akin to that experienced during the Great Depression is consistent with monetary policy (arguably with some fiscal help) having been able to limit the impact of the shock to the real economy. And perhaps the real effects of the shock could have been counteracted even more effectively had it not been for the limits imposed by the ZLB on nominal interest rates.[19] Therefore, the evidence might also be consistent with an improvement in the ability of policymakers to limit the damage caused by demand shocks on the real economy. This second conclusion, however, requires taking a stance on the size of the shock that hit the economy, in comparison, for instance, to the one that occurred in the early 1930s. Although many commentators have compared the extent of the financial disruption during the financial crisis to that associated with the Great Depression, it is difficult to make such a comparison formally. Therefore, we consider this second conclusion more tentative than the one regarding the reduction in the slope of the Phillips curve, which is more directly supported by the evidence.

An important caveat to the line of reasoning pursued in this section is that it is predicated on the assumption that EBP shocks, and the Great Recession, were both primarily demand disturbances. More precisely, the requirement is that they should affect inflation through their impact on the conventional measures of cost pressures that we have analyzed in section II. If, on the contrary, these disturbances reflect an important cost-push component, the evidence shown above is harder to interpret. Suppose that an increase in the EBP shock, caused by raising financing costs, induces firms to charge higher prices through channels that are not manifested in changes in the labor share or other standard measures of slack, and that this occurs to a greater extent in the second part of the sample than in the first. In this case, inflation might end up being flat in response to that shock, as in the post-1990 evidence above, since the impulse would act as a positive cost-push shock in the Phillips curve at the same time as it depresses other sources of cost pressure through the conventional demand channels.[20]

19. The extent to which the ZLB was binding during the Great Recession and its aftermath is much debated in the literature (Swanson and Williams 2014; Galí and Gambetti 2020; Eggertsson and Egiev 2020).

20. Christiano, Eichenbaum, and Trabandt (2015) and Gilchrist and others (2017) develop models of this kind.

IV.A. The Role of Inflation Expectations

Our SVAR analysis so far has focused on the direct relationship between inflation and unemployment, conditional on demand shocks. This conditional correlation is informative on the slope of the aggregate supply curve, helping to distinguish it from that of aggregate demand. However, to translate our findings on this conditional correlation into statements about the sensitivity of prices to costs, or equivalently into the slope of the expectations-augmented Phillips curve of Friedman and Phelps, we must take a stance on the dynamics of inflation expectations. The concern, as articulated by Stock (2011) and Barnichon and Mesters (2019b) for instance, is that anchored expectations produce a flatter aggregate supply curve, potentially confounding changes in the coefficient on real activity in the structural Phillips curve. In fact, many observers refer to the expectations-augmented Phillips curve in attributing the transition to stable inflation since the 1990s to a change in the dynamics of inflation expectations, rather than to a change in the responsiveness of prices to costs (Bernanke 2007; Blanchard 2016; Gordon 2018; Ball and Mazumder 2019).

One approach to addressing this concern is to be specific about the microfoundations of firms' pricing decision, as we do in the DSGE model of the next section. In this section, we take an intermediate step in evaluating the potential of a change in the dynamics of expectations to be behind the finding of a flatter Phillips curve after 1990. This step is based on the impulse responses to the EBP shock derived above. This approach is very similar to that pursued by Barnichon and Mesters (2019b), but in contrast to their claim, it confirms our main finding of a large decline in the slope of the structural Phillips curve over the past three decades. This result is also consistent with the DSGE analysis of section IV.

Consider the following Phillips curve,

$$\pi_t = \alpha \pi^e_{t+1|t} + (1-\alpha)\pi_{t-1} + \kappa(x_t + s_t), \tag{5}$$

which generalizes equation (2) by allowing for inflation inertia and potentially nonrational inflation expectations. This expression also imposes that inflation has no permanent effect on real activity, consistent with many empirical versions of the Phillips curve, which are vertical in the long run. By replacing the time index t with $t + h$, taking an expectation conditional on time-t information and a derivative with respect to a demand shock at time t on both sides, we obtain

$$R^{\pi}_h = \alpha R^{\pi^e}_h + (1-\alpha) R^{\pi}_{h-1} + \kappa R^{x}_h, \tag{6}$$

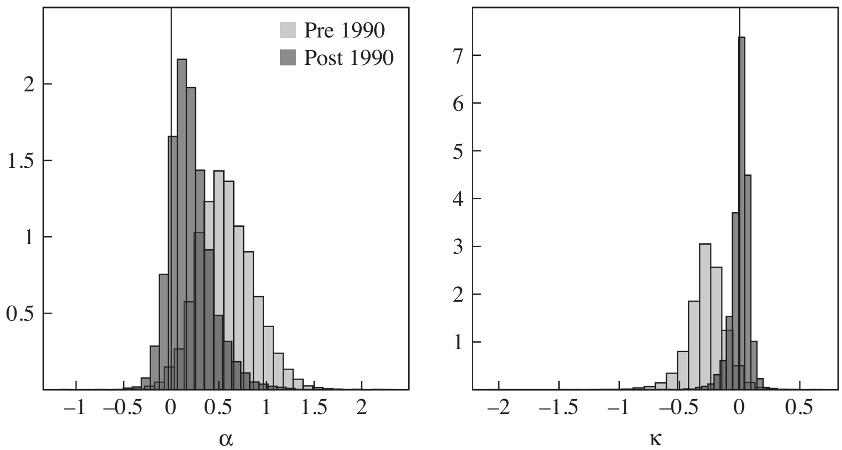
where $R_h^z = \frac{\partial E_t z_{t+h}}{\partial d_t}$ is the impulse response at horizon h of a generic variable z to the demand shock. In deriving equation (6), we used the fact that s_t is independent from the demand disturbance. Equation (6) says that, if equation (5) holds, the impulse responses of inflation, expected inflation, and the gap to a demand shock should satisfy equation (6) at every horizon h. Therefore, if we observed the true impulse responses of these variables, we would be able to infer the values of α and κ that satisfy equation (6).

In practice, the Phillips curve—equation (5)—does not hold exactly in the data. In addition, we can only estimate R_h^π, $R_h^{\pi^e}$, and R_h^x, as opposed to observing their population value. Despite these difficulties, we can still compute a value for α and κ by minimizing the distance between our estimates of the right- and the left-hand side of equation (6). To operationalize this strategy, we augment the structural VAR of the previous section with a measure of one-quarter-ahead inflation expectations from the Survey of Professional Forecasters. This VAR produces impulse responses that are nearly identical to those of figure 4. In addition, the reaction of expected inflation is similar to that of actual inflation, although more muted. For each posterior draw of the impulse responses of inflation, expected inflation, and unemployment (which proxies for the gap) to the EBP shock, we compute a value of α and κ using a linear projection. Figure 5 summarizes the outcome of this procedure. The right panel displays the histogram of the inferred value of κ in the pre- and post-1990 sample. This value clearly shifts toward zero in the second sample, consistent with the slope hypothesis. The estimate of α also declines, indicating a more inertial inflation process.

Accounting for inflation inertia in equation (5), which is important to fit the behavior of inflation as demonstrated by Fuhrer and Moore (1995), Galí and Gertler (1999), and Fuhrer (2010), is also crucial to the inference on the slope of the Phillips curve. If we restrict $\alpha = 1$, as in Barnichon and Mesters (2019b), our results are more consistent with theirs: κ still declines after 1990, but less than in figure 5.

Another difference from Barnichon and Mesters (2019b) is that their analysis is based on one-year-ahead inflation expectations from the Michigan Survey, as in Coibion and Gorodnichenko (2015). Unfortunately, this series is only available starting in 1978, which does not give us enough observations for a reliable estimation before 1990. However, figure 6 shows that the results for the post-1990 sample with the Michigan data are similar to our baseline with SPF data, confirming the robustness of our findings.

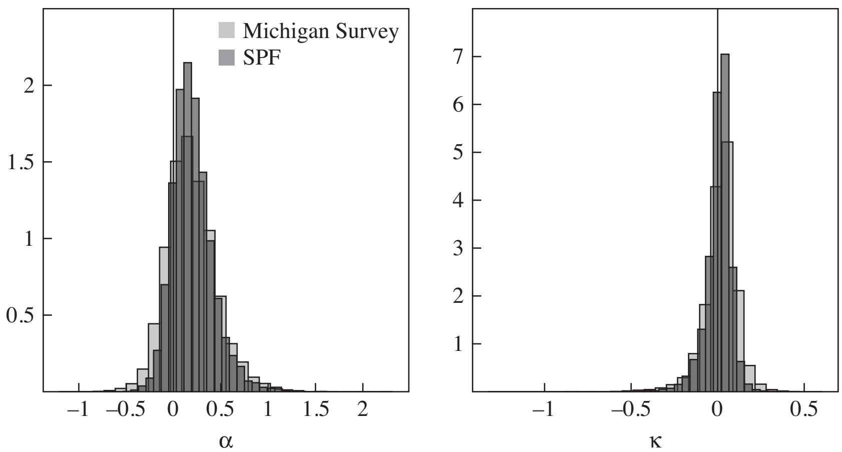
Figure 5. Posterior Distribution of Inferred Values of α and κ



Source: Authors' calculations.

Note:These values are obtained as described in section IV.A, using the impulse responses of inflation, expected inflation, and unemployment to an EBP shock, identified by assuming that it affects the EBP contemporaneously, but all other variables with a lag. The impulse responses are from the baseline VAR of section I.B, augmented with data on one-quarter-ahead inflation expectations and the EBP. The pre- and post-1990 samples consist of data from 1973:Q1 to 1989:Q4 and from 1989:Q1 to 2019:Q3, respectively.

Figure 6. Posterior Distribution of Inferred Values of α and κ in the Post-1990 Sample



Source: Authors' calculations.

Note: These values are obtained as described in section IV.A, using the impulse responses of inflation, expected inflation, and unemployment to an EBP shock, identified by assuming that it affects the EBP contemporaneously, but all other variables with a lag. The impulse responses are from the baseline VAR of section I.B, augmented with data on inflation expectations and the EBP. The Michigan Survey and SPF histograms are obtained using data on one-year-ahead inflation expectations from the Michigan Survey and one-quarter ahead inflation expectations from the Survey of Professional Forecasters, respectively.

V. Interpreting the Facts with an Estimated DSGE Model

As argued above, interpreting the facts of section II only through the lens of the VAR is challenging, as one can only speculate on the mechanisms behind them. For instance, the reduced sensitivity of inflation to labor market conditions in the second part of the sample could be due to differences in monetary policy or in other aspects of the economy. In the previous section, we made progress on this issue by identifying a demand disturbance in a structural VAR. In this section, instead, we interpret these facts through the lens of a fully specified DSGE model.

In such a structural model, interpreting the evidence is straightforward, since the mechanism is identified as part of the estimation. Of course, using structural models comes with its own challenges, most notably the fact that the model may be misspecified. From the perspective of the question at hand, such misspecification may imply that the DSGE cannot reproduce the same facts as the VAR. The first part of this section addresses this concern. It shows that the Federal Reserve Bank of New York's DSGE model, when estimated over the same subsamples as the VAR, reproduces the VAR facts both qualitatively and by and large quantitatively. In particular, a shock that has an impact on the labor market in ways comparable to those described in section II produces a muted response of inflation after 1990. The second part of the section investigates what changes in the model's structural parameters deliver these much attenuated responses. As in section IV, we focus on two explanations: changes in the conduct of monetary policy and changes in the slope of the structural price Phillips curve. We begin with a very brief description of the DSGE model.

V.A. The DSGE Model

The NY Fed DSGE model is a medium-scale New Keynesian model. In broad strokes, the model can be as Smets and Wouters (2007) plus financial frictions as in Bernanke, Gertler, and Gilchrist (1999) and Christiano, Motto, and Rostagno (2014). Variations of this model have been used in Del Negro and Schorfheide (2013), Del Negro, Giannoni, and Schorfheide (2015), and Del Negro and others (2017).[21] Cai and others (2019) document the model's real-time forecasting performance over the past ten years. The model is estimated using multiple measures of output growth (both

21. More specifically, the model is the same as that in Del Negro and others (2017), except that the processes for productivity growth and the safety premium were assumed to be more persistent in that paper.

GDP and gross domestic income), the growth in consumption, investment, the real wage, hours worked, two measures of inflation (both core PCE and the GDP deflator), long-run inflation expectations, the federal funds rate, the ten-year Treasury yield, the Baa-Treasury spread, and the series of total factor productivity growth constructed by Fernald (2012). We also allow for anticipated policy shocks as in Del Negro, Giannoni, and Patterson (2015) and Laséen and Svensson (2011), to account for the ZLB on nominal interest rates and forward guidance. Therefore, we augment the set of observables to incorporate interest rate expectations during the ZLB period. Online appendix B provides all the equilibrium conditions, the definition of the observables, and the specification of the priors, which are the same as in Smets and Wouters (2007) for all parameters common to the two models.

Here we only report two of the model's equations because their parameters are mentioned in the analysis that follows. The first equation is the price Phillips curve,

$$\pi_t = \kappa_p mc_t + \frac{\iota_p}{1+\iota_p\bar{\beta}}\pi_{t-1} + \frac{\bar{\beta}}{1+\iota_p\bar{\beta}}E_t[\pi_{t+1}] + \lambda_{f,t}, \tag{7}$$

where π_t is inflation, mc_t is marginal cost, and $\lambda_{f,t}$ captures exogenous fluctuations in desired markups. The slope of the Phillips curve is

$$\kappa_p = \frac{\left(1-\zeta_p\bar{\beta}\right)\left(1-\zeta_p\right)}{\left(1+\iota_p\bar{\beta}\right)\zeta_p\left(\left(\Phi_p-1\right)\epsilon_p+1\right)}.$$

Its key determinant is the Calvo parameter ζ_p, which represents the fraction of firms that do not adjust their price in every period. Therefore, a higher ζ_p means that prices are stickier, which makes the Phillips curve flatter. The other parameters entering the slope are the degree of indexation ι_p, the curvature parameter in the Kimball aggregator for prices ϵ_p, the size of fixed costs in production Φ_p, and $\bar{\beta}$, the discount rate adjusted for steady-state growth.

The second important equation is the interest rate rule followed by the monetary authority,

$$\begin{aligned} R_t = {} & \rho_R R_{t-1} + (1-\rho_R)\left(\psi_\pi(\pi_t - \pi_t^*) + \psi_y(y_t - y_t^*)\right) \\ & + \psi_{\Delta y}((y_t - y_t^*) - (y_{t-1} - y_{t-1}^*)) + r_t^m, \end{aligned} \tag{8}$$

where R_t is inflation, y_t and y_t^* are actual and natural output, so that their difference is a measure of the output gap, π_t^* is a time-varying inflation target, and r_t^m captures exogenous departures from the policy rule.[22]

V.B. Can the DSGE Replicate the VAR Facts?

To assess whether the estimated DSGE model can reproduce the VAR facts of section II, we compute the impulse responses to the U shock in the VAR implied by the estimated DSGE model. Del Negro and Schorfheide (2004) show how to construct the VAR approximation of the DSGE model, which they call DSGEVAR(∞). We refer to this approximation as DSGEVAR in the remainder of the paper.[23] The NY Fed DSGE model does not feature all the variables included in the VARs of section II, most notably unemployment. Therefore, the results in this section are based on a DSGEVAR with slightly fewer observables. These are hours worked per capita in log levels, a measure of price inflation, the labor share, and wage inflation.[24] This information is enough to capture the dynamics of the key economic variables involved in the connection between inflation and real activity.

Figure 7 shows the impulse responses of these variables to a U shock in two DSGEVARs, one based on the DSGEVAR implied by the pre-1990 DSGE estimates and the other reflecting the estimation on the post-1990 sample. The U shock is identified using the methodology described in section I.A., applied to hours rather than to unemployment. As in section II, the response of the labor market to a U shock is similar in the two samples, although it is slightly smaller on impact and more persistent in the second sample.[25] The response of the labor share is also similar in the two samples and, if anything, stronger in the post-1990 estimation. Qualitatively and

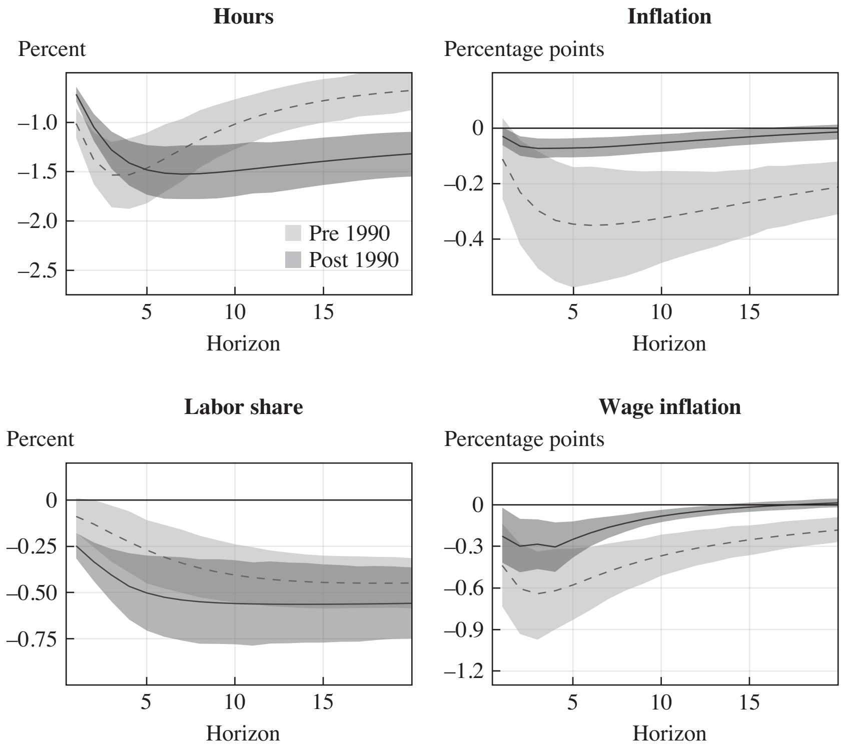
22. The time-varying inflation target π_t^*, which was not present in the original Smets and Wouters (2007) specification, is a very persistent process that captures the secular rise and fall of inflation and nominal interest rates in the estimation sample. As in Del Negro and Eusepi (2011), we use data on long-run inflation expectations as an observable in the estimation, which help the model to account for those low-frequency movements.

23. For any given vector of DSGE model parameters, the VAR approximation of the DSGE model is what would be obtained by generating artificial data from the DSGE model and estimating a VAR on such generated data. Del Negro and Schorfheide (2004) show how to compute the matrices of this VAR in population, without actually generating any artificial data.

24. The VARs used in section II also include the log level of per capita GDP. This variable is not stationary in the NY Fed DSGE model. Therefore we cannot include it in the DSGEVAR.

25. The DSGE model, and therefore the DSGEVAR, uses a slightly different measure of hours worked than the VAR in section IV.A. This is because the variable definitions in the NY Fed DSGE model are the same as in Smets and Wouters (2007).

Figure 7. Impulse Responses to an Unemployment Shock in the VAR Approximation of the DSGE



Source: Authors' calculations.

Note: The impulse responses are from the VAR approximation of the NY Fed DSGE model (DSGEVAR) described in section V. The shock is identified using a Cholesky strategy, with unemployment ordered first. The solid lines are posterior medians, while the shaded areas correspond to 95 percent posterior credible regions. The pre- and post-1990 samples consist of data from 1964:Q2 to 1989:Q4 and from 1990:Q1 to 2019:Q3, respectively.

quantitatively, these responses are also very similar to those in section II. The DSGE model therefore confirms that the transmission of U shocks to marginal costs is very similar across the two samples. On the contrary, the response of inflation is notably different: it is very muted in the second sample, as in the VAR. Finally, the responses of nominal wage inflation are somewhat weaker in the second sample, but not as weak as that of inflation, also consistent with the results of section II.

What changes in the estimated DSGE parameters are responsible for the differences in the DSGEVAR responses before and after 1990? To answer this question, we concentrate on the parameters that map more directly onto the alternative explanations of the muted response of inflation to the business cycle that we are comparing: the slope of the price and wage Phillips curve and the parameters describing monetary policy. The first row of figure 8 focuses on the former. The upper left panel shows the posterior distribution of the slope of the price Phillips curve κ_p in the two samples, which has declined substantially over time.[26] In contrast, the upper right panel of figure 8 shows that the posterior distribution of the slope of the wage Phillips curve (κ_w) is quite similar across samples, although it also shifts somewhat to the right.

The remaining two rows of figure 8 focus on the parameters of the monetary policy rule, which capture the inertia of the interest rate (ρ_R) and its response to inflation (ψ_π), the output gap level (ψ_y) and its growth ($\psi_{\Delta y}$). The posterior distribution of the response to inflation does not change much across samples. The response to the output gap increases after 1990, while that to its growth rate falls. Interest rate persistence is notably higher after 1990.[27] The implications of the estimated decline in the slope of the price Phillips curve κ_p for the dynamics of inflation are clear, at least qualitatively: they make inflation less responsive to real activity, as in the data. In contrast, the combined implications of the changes in the policy rule parameters are less obvious. Therefore, we now move to investigate which of these changes in estimated parameters might explain the observed changes in the response of inflation to U shocks.

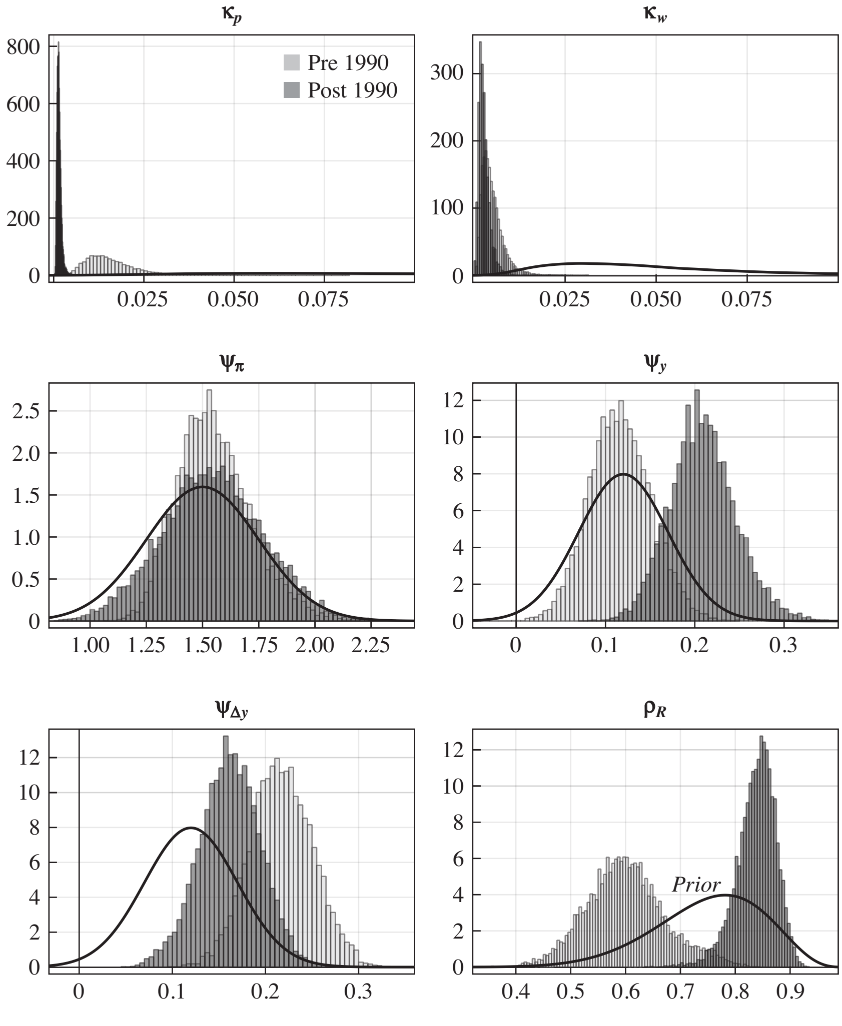
V.C. Explaining the Facts

Can the changes in the estimates of the slope of the price Phillips curve or the policy parameters quantitatively explain the facts of figure 7?

26. We focus on the slope of the price Phillips curve κ_p rather than on the underlying structural parameters for two reasons. First, the aggregate data used in the estimation of the DSGE model identify the slope, not the underlying structural parameters. For example, the price stickiness parameter ζ_p and Kimball aggregator parameter ϵ_p are not separately identified, as they only affect the slope and do not enter anywhere else in the model. We follow Smets and Wouters (2007) in estimating the former and calibrating the latter, but this is an arbitrary choice. Therefore, we cannot interpret the change in the slope reliably in terms of the underlying structural parameters. Second, only the slope, and not the underlying parameters, matters for the dynamics of the system and for the outcome of the counterfactuals presented below. With these caveats, online appendix figure C.1 reports the posterior estimates of the price stickiness parameter ζ_p, which, according to our estimation, is the key driver of the change in the slope.

27. During the recent ZLB episode, the model accounts for the forced deviations of the interest rate from the estimated policy rule, as well as for forward guidance, using anticipated monetary policy shocks.

Figure 8. Prior and Posterior Distributions for Selected Parameters of the DSGE Model



Source: Authors' calculations.
Note: The pre- and post-1990 samples consist of data from 1964:Q2 to 1989:Q4 and from 1990:Q1 to 2019:Q3, respectively.

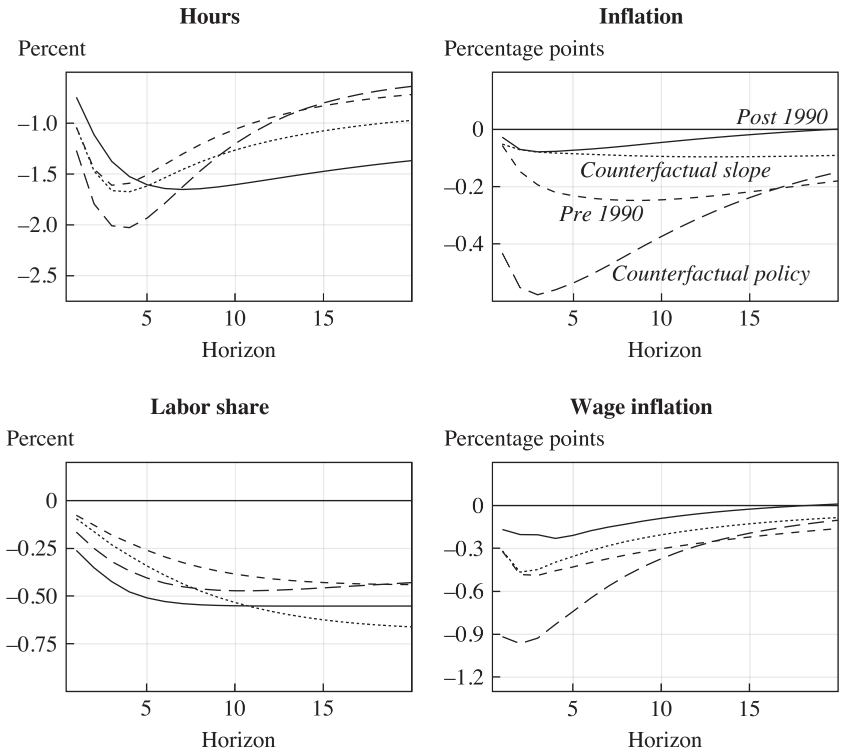
Figure 9 answers this question. The pre-1990 and post-1990 lines show the impulse responses to a business cycle shock in the DSGEVAR implied by the posterior mode of the DSGE parameters before and after 1990.[28] The other two lines show the counterfactual impulse responses obtained using the pre-1990 parameters, except for the slope of the Phillips curve and the policy-rule parameters. For these parameters, we use the posterior mode of the post-1990 estimates. The point of this exercise is the following: if changes in the slope (policy) fully account for the differences in the impulse responses across samples, then the counterfactual policy line should be as close as possible to the solid red line. In terms of the responses of hours and the labor share, both counterfactuals are close to the post-1990 line, which represents the actual post-1990 responses. However, this is not the case for the responses of price and wage inflation. For inflation, the slope counterfactual is essentially on top of the actual post-1990 response. In contrast, the policy counterfactual produces an even stronger reaction than before 1990. Finally, both counterfactuals tend to overestimate the response of wage inflation, with the policy counterfactual again faring worse than the slope one.

Figure 9 indicates that changes in the slope of the price Phillips curve alone can explain the muted response of inflation to U shocks. Together with the fact that the policy rule is not very different before and after 1990, except for the estimate of ρ_R, this result suggests that a change in the policy rule is not necessary to account for the behavior of inflation after 1990. But could there still be a role for policy if we took a change in the slope of the Phillips curve off the table?

To answer this question, we estimate the DSGE model before and after 1990 allowing only the policy rule parameters to change. The other parameters are assumed to be constant over the entire sample. The purpose of this exercise is to give the policy hypothesis the best shot at explaining the facts. The policy rule parameters estimated as part of this exercise are reported in figure 10. They are different from those in figure 8, which are based on an estimation in which all parameters can change before and after 1990 to best fit the data. First, the distribution of the response to inflation (ψ_π) is skewed to the right after 1990, relative

28. These responses are essentially the same as those in figure 7. The difference is that, there, the solid lines are the medians of the posterior distribution of the impulse responses. Here, the solid lines are the impulse responses at the mode of the posterior distribution of the parameters. We use the posterior mode of the parameters here because this approach makes it easier to conduct the counterfactual described below.

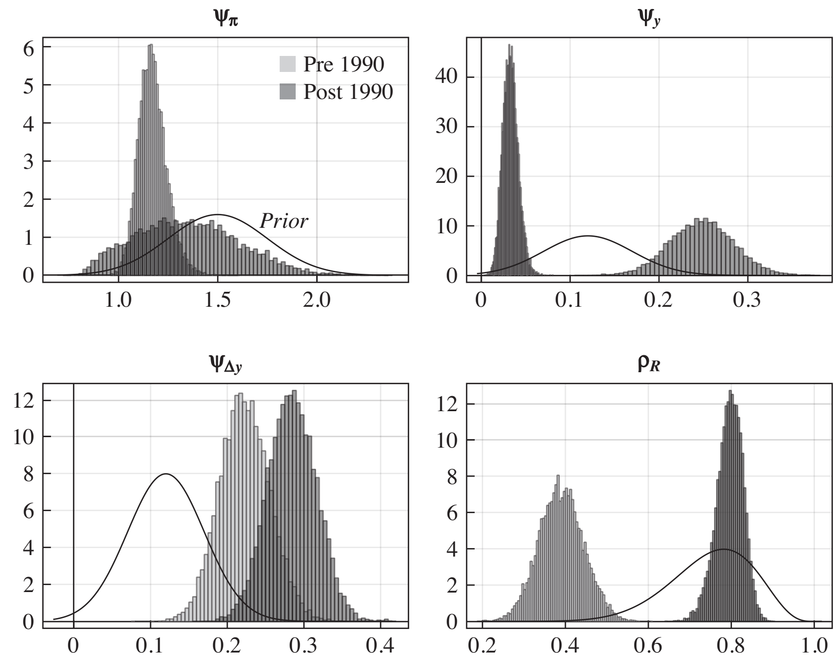
Figure 9. Impulse Responses to an Unemployment Shock in the DSGEVAR, with Counterfactuals



Source: Authors' calculations.

Note: The impulse responses are from the VAR approximation of the NY Fed DSGE model (DSGEVAR) described in section V. The shock is identified using a Cholesky strategy, with unemployment ordered first. The pre-1990 and post-1990 lines are impulse responses constructed using the modal posterior estimates in the pre- and post-1990 samples, respectively. The counterfactual slope responses are obtained using the pre-1990 modal posterior estimates for all parameters except the slope of the Phillips curve, for which we use the post-1990 posterior mode. The counterfactual policy responses are obtained using the pre-1990 modal posterior estimates for all parameters except the policy rule ones, for which we use the post-1990 posterior mode.

Figure 10. Prior and Posterior Distributions for Selected Parameters of the DSGE Model, Estimated with Different Monetary Policy before and after 1990

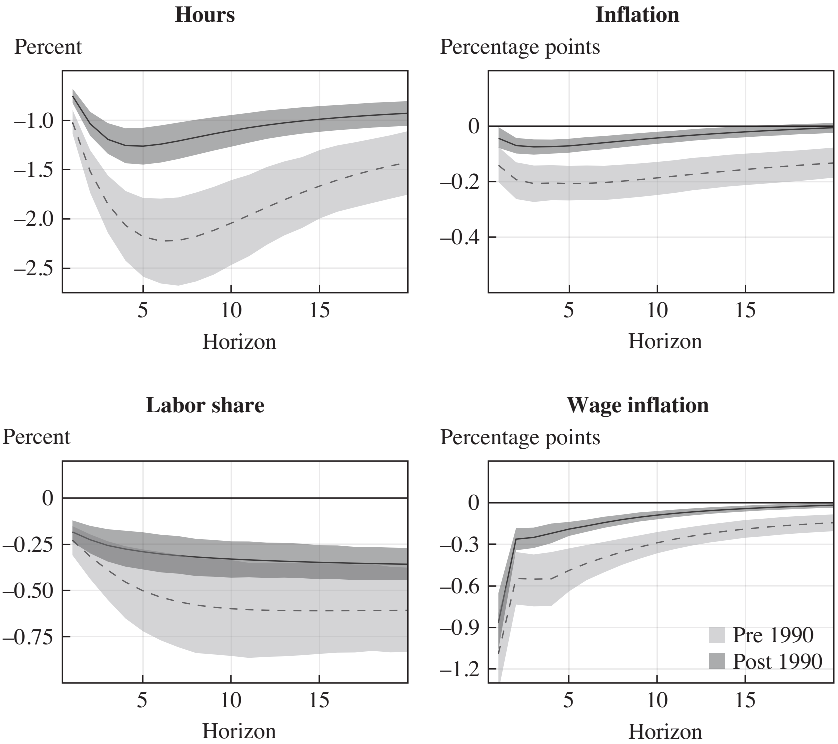


Source: Authors' calculations.
Note: In this experiment, the model is estimated allowing the coefficients of the monetary policy rule to be different before and after 1990, but assuming that all other coefficients are constant for the entire sample from 1964:Q2 to 2019:Q3.

to the pre-1990 estimates. Second, the differences in the response to the output gap (ψ_y) before and after 1990 are larger than in figure 8. Third, the response to the output gap growth ($\psi_{\Delta y}$) is now also stronger in the second sample, rather than weaker. Finally, the persistence in interest rates (ρ_R) remains much higher post 1990, which implies that monetary policy has stronger control of inflation, as shown for instance by Woodford (2003).

Together, these changes in the policy rule parameters do result in a more muted response of inflation to the U shock in the second sample, as shown in figure 11. This is not too surprising, since this version of the model must account for the stability of inflation after 1990 using only changes in the policy parameters. The conclusion is that it is possible to find changes in

Figure 11. Impulse Responses to an Unemployment Shock in the DSGEVAR, Estimated with Different Monetary Policy before and after 1990



Source: Authors' calculations.

Note: The impulse responses are from the VAR approximation of the NY Fed DSGE model (DSGEVAR) described in section V. The shock is identified using a Cholesky strategy, with unemployment ordered first. The solid lines are posterior medians, while the shaded areas correspond to 95 percent posterior credible regions. In this experiment, the model is estimated allowing the coefficients of the monetary policy rule to be different before and after 1990, but assuming that all other coefficients are constant for the entire sample from 1964:Q2 to 2019:Q3.

the policy rule parameters that stabilize inflation as in the data. However, this success in the inflation dimension comes at the cost of also muting the responses of hours and the labor share after 1990. This is in contrast with the empirical facts highlighted in section II and in figure 7, in which hours and the labor share are, if anything, more responsive after 1990. These results can be interpreted in light of the discussion in section III: to the extent that the main shocks hitting the economy are demand shocks, policy can effectively control inflation by neutralizing their real effects. But the evidence that this has occurred is weak.[29]

These results stand in contrast to those of figure 12, where we repeat the same exercise imposing that only the slope of the price Phillips curve has changed between the two samples. The responses in figure 12 are in line with those in figure 7: the reaction of hours and the labor share is very similar across samples, but that of inflation is much more muted after 1990.[30]

VI. Policy Implications

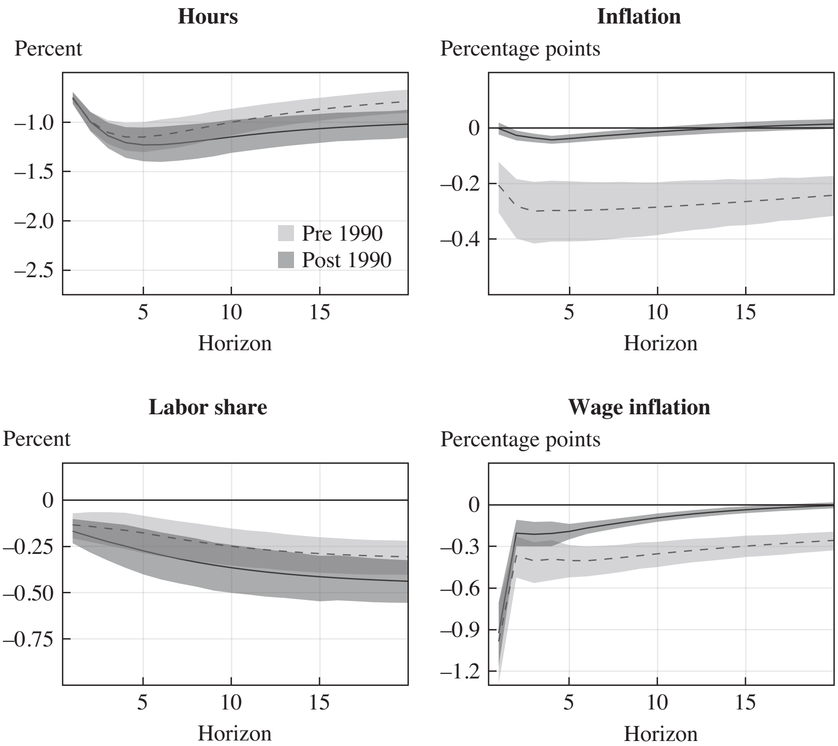
To understand the policy implications of our findings, a useful starting point is the simple model of section III. In that framework, both real activity and inflation increase in response to a positive shift in aggregate demand. With a flatter aggregate supply curve, however, any given shift in demand has a larger impact on real activity and a smaller effect on inflation. As it turns out, this simple "static" intuition also holds in the quantitative DSGE model of section V: in response to a monetary policy shock, either a short-lived or a very persistent one, inflation reacts less and real activity reacts more when the Phillips curve slope is low. In essence, with a flatter Phillips curve, it is more difficult for monetary authorities to steer inflation in any particular direction using unsystematic policy measures.

The fact that monetary policy cannot as easily control inflation by engineering isolated, unexpected shifts in aggregate demand should not be surprising. After all, a flatter Phillips curve corresponds to a structural change in the economy. As a consequence, the best way for monetary policy to reestablish its control of inflation should be to change its systematic

29. The exercise that we just described might not in fact be the best shot for the policy hypothesis because the post-1990 sample includes a long period at the ZLB, when monetary policy could not react as much to U shocks. Figures C.2 and C.3 in the online appendix reproduce the results in figures 10 and 11 using a sample ending in 2008:Q3, before the onset of the ZLB period and the Great Recession. These parameter estimates indicate an even stronger response to inflation in the post-1990 period, but the impulse responses in figure C.3 do not suggest any attenuation in the response of inflation to a U shock after 1990, contrary to the evidence.

30. The corresponding posterior distributions of the slope of the price Phillips curve κ_p pre- and post-1990 are shown in figure C.4 in the online appendix. They are very similar to those in figure 8. In this exercise the slope of the price Phillips curve changes across samples due to changes in the price stickiness parameter ζ_p. Results are similar when both the slope of the price Phillips curve κ_p and the indexation parameter ι_p are allowed to change.

Figure 12. Impulse Responses to an Unemployment Shock in the DSGEVAR, Estimated with Different Slopes of the Price Phillips Curve before and after 1990



Source: Authors' calculations.

Note: The impulse responses are from the VAR approximation of the NY Fed DSGE model (DSGEVAR) described in section V. The shock is identified using a Cholesky strategy, with unemployment ordered first. The solid lines are posterior medians, while the shaded areas correspond to 95 percent posterior credible regions. In this experiment, the model is estimated allowing the slope of the price Phillips curve to be different before and after 1990, but assuming that all other coefficients are constant for the entire sample from 1964:Q2 to 2019:Q3.

reaction to the state of the economy, as opposed to through a series of policy shocks. In addition, intuitively, the value of adopting such a policy rule should be higher exactly when the slope of the Phillips curve is low and the trade-off implied by a business-as-usual policy is unfavorable.

We illustrate this principle quantitatively using the DSGE model discussed in the previous section. We begin by showing that our main explanation for the muted response of inflation to real activity—a lower slope of the Phillips curve—is also consistent with the persistent weakness of inflation since the Great Recession. We then demonstrate that alternative policy rules—such as average inflation targeting—can be as effective at stabilizing inflation with a flat Phillips curve as it would have been with a steeper one.

Figure 13 shows the model's forecasts of inflation and marginal costs starting in 2020:Q1 under the estimated policy rule and two alternative settings for the slope of the price Phillips curve κ_p: post-1990 and pre-1990. As shown in figure 8 both the pre- and post-1990 values of κ_p are quite low in absolute terms. For ease of exposition, and to emphasize the relative differences, we will refer to the pre-1990 κ_p as "high" and to the post-1990 κ_p as "low" in the remainder of the section.[31]

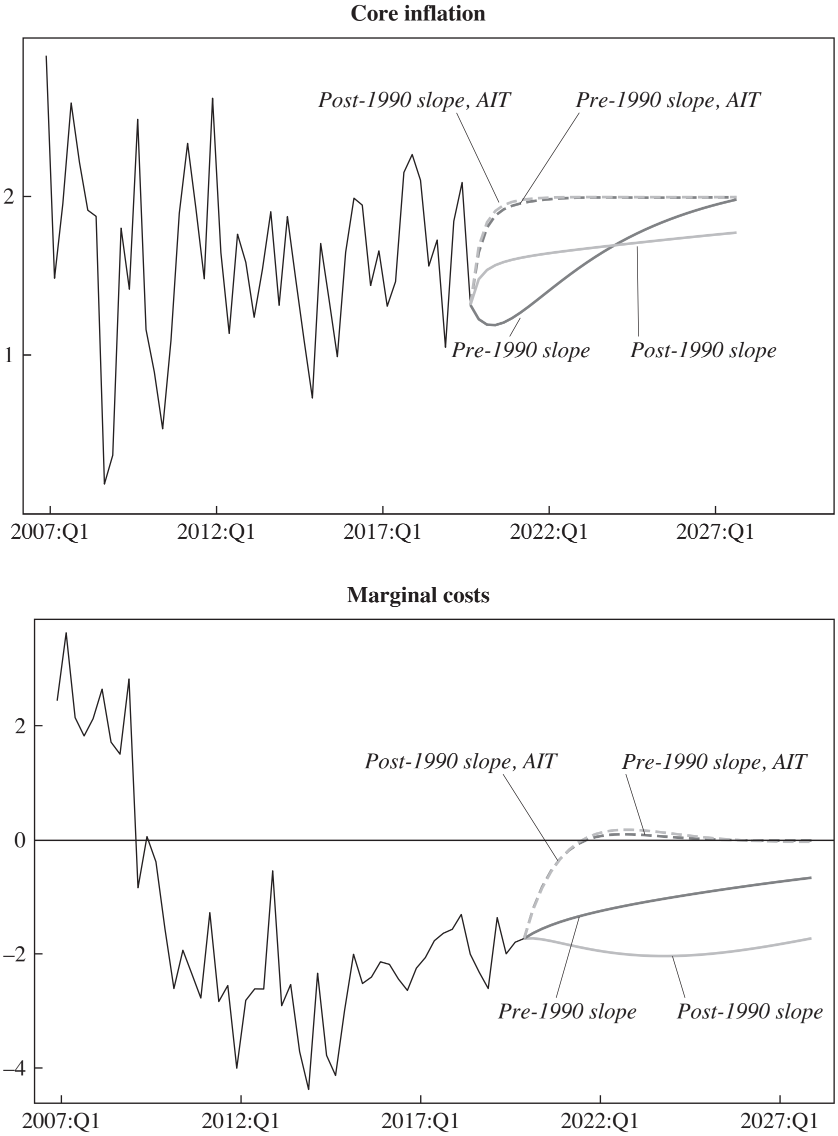
When κ_p is high, inflation falls more initially, but then it reverts relatively quickly to the Federal Open Market Committee's long-run goal. When κ_p is low, the shortfall of inflation is much more persistent, to the point that it is not closed even by the end of the forecast horizon (left panel). The differences in the inflation forecasts can be explained in terms of those for real marginal costs (right panel). When κ_p is high, marginal costs revert more quickly to the steady state, while they remain persistently depressed with a low κ_p. Intuitively, a high κ_p brings the economy closer to one with flexible prices, in which real marginal costs never deviate from their desired level. With a low κ_p, on the contrary, the deviations of marginal costs from their desired level is more persistent, due to a stronger endogenous propagation.[32] In the current situation, this means that the drag on marginal costs brought about by the Great Recession and by the other negative shocks that followed continues to exert a negative impact on inflation. Therefore, a flat estimated Phillips curve in the NY Fed DSGE model contributes toward explaining why inflation has been persistently below target over the past decade.

We now repeat the same exercise, but under the alternative policy rule of average inflation targeting (AIT). We focus on this specific rule because it is a much discussed monetary policy strategy in the current

31. The two values for κ_p, as well as the values of the other DSGE model parameters used to generate the forecasts, are the modal estimates of the two-regime estimation where only the slope of the price Phillips curve is allowed to be different across the two regimes. The forecast is computed under the assumption that κ_p returns to its pre-1990 value starting in 2020:Q1.

32. In his discussion, Olivier Blanchard makes the point that markups, which are the inverse of the real marginal costs shown in figure 13, are more countercyclical in recent decades than they had been before. His evidence is consistent with figure 13: after 1990 the economy has become more Keynesian and real marginal costs fall more, and more persistently, following a recession.

Figure 13. DSGE Forecasts of Core PCE Inflation and Marginal Costs under Different Monetary Policies



Source: Authors' calculations.

Note: These forecasts use the modal estimate of κ_p in the pre- and post-1990 period, under the estimated policy rule, and a counterfactual policy of average inflation targeting. The lines going back to 2007 represent data for inflation and the smoothed values of marginal costs, which are not observed. The two values for κ_p, as well as the values of the other DSGE parameters used to generate the forecasts, are the modal estimates of the two-regime estimation where only the slope of the price Phillips curve is allowed to be different across the pre- and post-1990 regimes.

debate (Mertens and Williams 2019). The exact AIT rule used in the counterfactual simulations is

$$R_t = \frac{0.25}{1-\gamma}\pi gap_t, \tag{9}$$

where the inflation gap is computed as

$$\pi gap_t = \pi_t + \gamma \pi gap_{t-1}$$

and all the variables are in deviation from their steady state. For inflation, this is 2 percent—the Federal Open Market Committee's long-run goal.[33]

The dashed lines in figure 13 show that switching to AIT has a more significant effect on the dynamics of marginal costs in the low κ_p economy: the gap between the solid and dashed lines is larger and more persistent with the low post-1990 slope than with the high pre-1990 one. As a consequence, marginal costs quickly revert to steady state, leading to nearly identical paths of inflation in the high and low κ_p economies.

In sum, a flat Phillips curve requires the monetary authority to work harder to stabilize inflation: unemployment needs to get lower to bring inflation back to target after a recession, everything else being equal. In equilibrium, however, a flat Phillips curve also makes the economy more Keynesian, implying that systematic monetary policy can persistently affect the dynamics of marginal costs (Del Negro, Giannoni, and Schorfheide 2015). As a result, the ability of policy to achieve its objectives is not compromised. A corollary of this general principle is that systematic policies like average inflation targeting could be especially effective in bringing inflation back to 2 percent in the current environment, as shown by the simulations above. An important caveat to this sanguine conclusion is that we obtained it in a rational expectations New Keynesian model where private agents perfectly understand the monetary policy strategy in place.[34]

33. The AIT gap πgap_t is the discounted average of past inflation rates, where we set $\gamma = 0.93$ to produce a half-life of ten quarters. In this example, we set $\pi gap_t = 0$ at the beginning of the forecast, resulting in no overshooting. Lower initial values of the gap, reflecting the accumulated shortfall of inflation from target over the past few years, would generate an overshoot of inflation over 2 percent.

34. One relevant feature of this class of models is the so-called forward guidance puzzle (Del Negro, Giannoni, and Patterson 2015), according to which policies based on promises of future actions are too successful in stabilizing the economy. This is one of the channels through which AIT works well to stabilize inflation in our simulations. For an assessment of the effectiveness of AIT under bounded rationality, see Budianto, Nakata, and Schmidt (2020).

However, the idea that monetary policy has a tighter grip on the real economy when aggregate supply is flatter goes beyond the confines of this specific class of models. We conclude that, even in the current environment, monetary policy can be as effective as it ever was in achieving price stability, as long as it pursues the appropriate strategy.

VII. Concluding Remarks

How do inflation and unemployment co-move over the business cycle? This paper explored how the answer to this question changed over the past few decades, and why. Unlike much of the existing Phillips curve literature, we address this question in a multivariate, dynamic context—studying impulse responses to typical unemployment shocks in a VAR, rather than regression coefficients in a Phillips curve—because the persistence of the business cycle impulses matters, and because looking jointly at several variables provides more clues on possible mechanisms and explanations.

We find that the persistence of unemployment fluctuations has increased somewhat after 1990, consistent with longer but shallower recoveries. The same is true of many other measures of labor and goods market activity that are commonly employed as proxies for cost pressures, including real wages and unit labor costs. The exception to this picture of relative stability over time is inflation, which has become far less sensitive to business cycle shocks.

Who is the culprit? Together, our findings rule out explanations that hinge on unemployment having become less relevant as an indicator of wage and price pressures or on the demise of the wage Phillips curve. We are left to consider two main possibilities, which have potentially very different implications for monetary policy. One is a lower slope of the price Phillips curve, leading to a more Keynesian economy, in which demand shocks dominate business cycles. The other is that policy is better able and willing to stabilize inflation, making the economy more neoclassical, with fluctuations dominated by supply shocks instead.

The casual observation that inflation and unemployment were on the same side of the Federal Reserve's dual mandate for many years after the Great Recession casts some doubt on the policy hypothesis. To support this observation more formally, we use a structural VAR and the NY Fed DSGE model to identify shocks and mechanisms. The SVAR suggests that inflation has become less responsive to shocks to credit spreads that have a large impact on real activity. If these shocks mostly shift aggregate demand, this evidence supports the slope hypothesis. This is the same conclusion reached by the DSGE model. Although it imposes more restrictions than the VAR, the DSGE can replicate the VAR impulse responses, and it attributes the reduction in the sensitivity of inflation to business cycle shocks to a reduction in the slope of the structural Phillips curve. Changes in the policy parameters, alone, are less successful at explaining the facts.

Although our analysis points more decisively in the direction of the slope hypothesis than of the policy hypothesis, it does not imply that monetary policy did not play a role in stabilizing inflation. By most accounts, it was the Federal Reserve under Volcker that brought inflation under control in the early 1980s, when our estimates find that the price Phillips curve was still alive and well. Moreover, our study leaves a number of important questions unanswered. First among them are what structural forces underlie the reduced sensitivity of inflation to cost pressures. We leave this question for future research.

ACKNOWLEDGMENTS Prepared for the 2020 Spring Brookings Papers on Economic Activity. We thank Ethan Matlin and especially William Chen for excellent research assistance, our discussants Olivier Blanchard and Chris Sims, as well as Richard Crump, Keshav Dogra, Simon Gilchrist, Pierre-Olivier Gourinchas, Rick Mishkin, Argia Sbordone, and the conference participants for useful comments and suggestions. Jim Stock provided much appreciated guidance during the research and writing process. The views expressed in this paper are those of the authors and do not necessarily represent those of the European Central Bank, the Eurosystem, the Federal Reserve Bank of New York, or the Federal Reserve System.

References

Abraham, Katherine G., John C. Haltiwanger, and Lea E. Rendell. 2020. "How Tight Is the US Labor Market?" In the present volume of *Brookings Papers on Economic Activity*.

Afrouzi, Hassan, and Choongryul Yang. 2020. "Dynamic Rational Inattention and the Phillips Curve." Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3465793#.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas. 2019. "Business Cycle Anatomy." Working Paper. https://economics.mit.edu/files/15230.

Atkeson, Andrew, and Lee E. Ohanian. 2001. "Are Phillips Curves Useful for Forecasting Inflation?" *Federal Reserve of Minneapolis Quarterly Review* 25, no. 1: 2–11.

Auer, Raphael, and Andreas M. Fischer. 2010. "The Effect of Low-Wage Import Competition on U.S. Inflationary Pressure." *Journal of Monetary Economics* 57, no. 4: 491–503.

Ball, Laurence, and Sandeep Mazumder. 2011. "Inflation Dynamics and the Great Recession." *Brookings Papers on Economic Activity*, Spring, 337–405.

Ball, Laurence, and Sandeep Mazumder. 2019. "A Phillips Curve with Anchored Expectations and Short-Term Unemployment." *Journal of Money, Credit and Banking* 51, no.1: 111–37.

Bańbura, Marta, Domenico Giannone, and Michele Lenza. 2015. "Conditional Forecasts and Scenario Analysis with Vector Autoregressions for Large Cross-Sections." *International Journal of Forecasting* 31, no. 3: 739–56.

Barnichon, Régis, and Geert Mesters. 2019a. "Identifying Modern Macro Equations with Old Shocks," Discussion Paper 13765. London: Centre for Economic Policy Research. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=13765.

Barnichon, Régis, and Geert Mesters. 2019b. "The Phillips Multiplier." Discussion Paper 13480. London: Centre for Economic Policy Research. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=13480.

Bernanke, Ben S. 2007. "Inflation Expectations and Inflation Forecasting." Speech given at the Monetary Economics Workshop of the National Bureau of Economic Research Summer Institute, Cambridge, Mass., July 10.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." In *Handbook of Macroeconomics*, vol. 1C, edited by John B. Taylor and Michael Woodford. Amsterdam: North-Holland.

Blanchard, Olivier. 2016. "The Phillips Curve: Back to the '60s?" *American Economic Review* 106, no. 5: 31–34.

Blanchard, Olivier, Eugenio Cerutti, and Lawrence Summers. 2015. "Inflation and Activity—Two Explorations and Their Monetary Policy Implications." Working Paper 21726. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w21726.

Budianto, Flora, Taisuke Nakata, and Sebastian Schmidt. 2020. "Average Inflation Targeting and the Interest Rate Lower Bound." Working Paper 852. Basel: Bank for International Settlements. https://www.bis.org/publ/work852.htm.

Cai, Michael, Marco Del Negro, Marc P. Giannoni, Abhi Gupta, Pearl Li, and Erica Moszkowski. 2019. "DSGE Forecasts of the Lost Recovery." *International Journal of Forecasting* 35, no. 4: 1770–89.

Caldara, Dario, Cristina Fuentes-Albero, Simon Gilchrist, and Egon Zakrajšek. 2016. "The Macroeconomic Impact of Financial and Uncertainty Shocks." *European Economic Review* 88, 185–207.

Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12, no. 1: 383–98.

Carvalho, Carlos, Stefano Eusepi, Emanuel Moench, and Bruce Preston. 2019. "Anchored Inflation Expectations." Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3018198.

Cavallo, Alberto. 2018. "More Amazon Effects: Online Competition and Pricing Behaviors." Working Paper 25138. Cambridge, Mass.: National Bureau of Economic Research.

Christiano, Lawrence J., Martin S. Eichenbaum, and Mathias Trabandt. 2015. "Understanding the Great Recession." *American Economic Journal: Macroeconomics* 7, no. 1: 110–67.

Christiano, Lawrence J., R. Motto, and M. Rostagno. 2014. "Risk Shocks." *American Economic Review* 104, no. 1: 27–65.

Clark, Todd E., and Taeyoung Doh. 2014. "Evaluating Alternative Models of Trend Inflation." *International Journal of Forecasting* 30, no. 3: 426–48.

Coibion, Olivier, and Yuriy Gorodnichenko. 2015. "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation." *American Economic Journal: Macroeconomics* 7, no. 1: 197–232.

Coibion, Olivier, Yuriy Gorodnichenko, and Dmitri Koustas. 2013. "Amerisclerosis? The Puzzle of Rising U.S. Unemployment Persistence." *Brookings Papers on Economic Activity*, Fall, 193–260.

Crump, Richard K., Stefano Eusepi, Marc P. Giannoni, and Ayşegül Şahin. 2019. "A Unified Approach to Measuring *u**." *Brookings Papers on Economic Activity*, Spring, 143–238.

Daly, Mary C., and Bart Hobijn. 2014. "Downward Nominal Wage Rigidities Bend the Phillips Curve." *Journal of Money, Credit and Banking* 46, no. S2: 51–93.

Davis, Steven J. 2019. Comment on "A Unified Approach to Measuring *u**." *Brookings Papers on Economic Activity*, Spring, 223–30.

Del Negro, Marco, and Stefano Eusepi. 2011. "Fitting Observed Inflation Expectations." *Journal of Economic Dynamics and Control* 35, no. 12: 2105–31.

Del Negro, Marco, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti. 2017. "Safety, Liquidity, and the Natural Rate of Interest." *Brookings Papers on Economic Activity*, Spring, 235–316.

Del Negro, Marco, Marc P. Giannoni, and Christina Patterson. 2015. "The Forward Guidance Puzzle." Staff Report 574. Federal Reserve Bank of

New York. https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr574.pdf.

Del Negro, Marco, Marc P. Giannoni, and Frank Schorfheide. 2015. "Inflation in the Great Recession and New Keynesian Models." *American Economic Journal: Macroeconomics* 7, no. 1: 168–96.

Del Negro, Marco, and Frank Schorfheide. 2004. "Priors from General Equilibrium Models for VARs." *International Economic Review* 45, no. 2: 643–73.

Del Negro, Marco, and Frank Schorfheide. 2013. "DSGE Model-Based Forecasting." In *Handbook of Economic Forecasting*, vol. 2, edited by Graham Elliott and Allan Timmermann. Amsterdam: North-Holland.

Eggertsson, Gauti B., and Sergey K. Egiev. 2020. "Fundamental Driven Liquidity Traps: A Unified Theory of the Great Depression and the Great Recession." Working Paper. https://drive.google.com/file/d/1vKfVvF9RwsMiMoqdEdO-wfbrdLbuM345/view.

Faccini, Renato, and Leonardo Melosi. 2020. "Bad Jobs and Low Inflation." Working Paper 2020-09. Federal Reserve Bank of Chicago. https://www.chicagofed.org/publications/working-papers/2020/2020-09.

Fernald, John. 2012. "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity." Working Paper 2012-19. Federal Reserve Bank of San Francisco. http://www.frbsf.org/economic-research/publications/working-papers/2012/19/.

Fitzgerald, Terry J., and Juan Pablo Nicolini. 2014. "Is There a Stable Relationship between Unemployment and Future Inflation? Evidence from U.S. Cities." Working Paper 713. Federal Reserve Bank of Minneapolis. https://www.minneapolisfed.org/research/working-papers/is-there-a-stable-relationship-between-unemployment-and-future-inflation-evidence-from-us-cities.

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

Forbes, Kristin J. 2019b. "How Have Shanghai, Saudi Arabia, and Supply Chains Affected U.S. Inflation Dynamics?" *Review* 101, no. 1: 27–44.

Forbes, Kristin J., Joseph E. Gagnon, and Christopher G. Collins. 2020. "Low Inflation Bends the Phillips Curve around the World." Working Paper 20-06. Washington: Peterson Institute for International Economics. https://www.piie.com/publications/working-papers/low-inflation-bends-phillips-curve-around-world.

Fuhrer, Jeffrey C. 2010. "Inflation Persistence." In *Handbook of Monetary Economics*, edited by Benjamin M. Friedman and Michael Woodford. Amsterdam: North-Holland.

Fuhrer, Jeffrey C., and George Moore. 1995. "Inflation Persistence." *Quarterly Journal of Economics* 110, no. 1: 127–59.

Galí, Jordi. 2015. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*, 2nd ed. Princeton, N.J.: Princeton University Press.

Galí, Jordi, and Luca Gambetti. 2019. "Has the U.S. Wage Phillips Curve Flattened? A Semi-Structural Exploration." Working Paper 25476. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25476.

Galí, Jordi, and Luca Gambetti. 2020. "On the Empirical (Ir)relevance of the Zero Lower Bound Constraint." in *NBER Macroeconomics Annual* 34:141–70.

Galí, Jordi, and Mark Gertler. 1999. "Inflation Dynamics: A Structural Econometric Analysis." *Journal of Monetary Economics* 44, no. 2: 195–222.

Geerolf, François. 2020. "The Phillips Curve: A Relation between Real Exchange Rate Growth and Unemployment." Working Paper. https://fgeerolf.com/phillips.pdf.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri. 2015. "Prior Selection for Vector Autoregressions." *Review of Economics and Statistics* 97, no. 2: 436–51.

Giannone, Domenico, Michele Lenza, and Lucrezia Reichlin. 2019. "Money, Credit, Monetary Policy, and the Business Cycle in the Euro Area: What Has Changed since the Crisis?" *International Journal of Central Banking* 15, no. 5: 137–73.

Gilchrist, Simon, Raphael Schoenle, Jae Sim, and Egon Zakrajšek. 2017. "Inflation Dynamics during the Financial Crisis." *American Economic Review* 107, no. 3: 785–823.

Gilchrist, Simon, Vladimir Yankov, and Egon Zakrajšek. 2009. "Credit Market Shocks and Economic Fluctuations: Evidence from Corporate Bond and Stock Markets." *Journal of Monetary Economics* 56, no. 4: 471–93.

Gilchrist, Simon, and Egon Zakrajšek. 2012. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review* 102, no. 4: 1692–720.

Goldfeld, Stephen M., and Alan S. Blinder. 1972. "Some Implications of Endogenous Stabilization Policy." *Brookings Papers on Economic Activity*, no. 3: 585–644.

Goolsbee, Austan D., and Peter J. Klenow. 2018. "Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce." In *American Economic Association Papers and Proceedings*, vol. 108. Pittsburgh, Pa.: AEA Publications.

Gordon, Robert J. 2013. "The Phillips Curve Is Alive and Well: Inflation and the NAIRU during the Slow Recovery." Working Paper 19390. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w19390.

Gordon, Robert J. 2018. "Friedman and Phelps on the Phillips Curve Viewed from a Half Century's Perspective." *Review of Keynesian Economics* 6, no. 4: 425–36.

Hall, Robert E. 2011. "The Long Slump." *American Economic Review* 101, no. 2: 431–69.

Hazell, Jonathan, Juan Herreno, Emi Nakamura, and Jón Steinsson. 2020. "The Slope of the Phillips Curve: Evidence from US States." Working Paper.

Hong, Gee Hee, Zsóka Kóczán, Weicheng Lian, and Malhar Nabar. 2018. "More Slack than Meets the Eye? Recent Wage Dynamics in Advanced Economies." Working Paper 18/50. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2018/03/09/More-Slack-than-Meets-the-Eye-Recent-Wage-Dynamics-in-Advanced-Economies-45692.

Hooper, Peter, Frederic S. Mishkin, and Amir Sufi. 2020. "Prospects for Inflation in a High Pressure Economy: Is the Phillips Curve Dead or Is It Just Hibernating?" *Research in Economics* 74, no. 1: 26–62.

Jones, Charles I. 2018. *Macroeconomics*, 4th ed. New York: W.W. Norton.

Jorgensen, Peter Lihn, and Kevin J. Lansing. 2019. "Anchored Inflation Expectations and the Flatter Phillips Curve." Working Paper 2019-27. Federal Reserve Bank of San Francisco. https://www.frbsf.org/economic-research/publications/working-papers/2019/27/.

Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti. 2010. "Investment Shocks and Business Cycles." *Journal of Monetary Economics* 57, no. 2: 132–45.

Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti. 2011. "Investment Shocks and the Relative Price of Investment." *Review of Economic Dynamics* 14, no. 1: 101–21.

Kareken, John, and Robert M. Solow. 1963. "Lags in Fiscal and Monetary Policy." In *Stabilization Policies: A Series of Research Studies Prepared for the Commission on Money and Credit*, edited by E. Cary Brown. Englewood Cliffs, N.J.: Prentice Hall.

Laséen, Stefan, and Lars E. O. Svensson. 2011. "Anticipated Alternative Policy-Rate Paths in Policy Simulations." *International Journal of Central Banking* 7, no. 3: 1–35.

Lindé, Jesper, and Mathias Trabandt. 2019. "Resolving the Missing Deflation Puzzle." Discussion Paper 13690. London: Centre for Economic Policy Research. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=13690.

Mackowiak, Bartosz, and Mirko Wiederholt. 2009. "Optimal Sticky Prices under Rational Inattention." *American Economic Review* 99, no. 3: 769–803.

Mankiw, N. Gregory, and Ricardo Reis. 2002. "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *Quarterly Journal of Economics* 117, no. 4: 1295–328.

Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H. Stock. 2014. "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve." *Journal of Economic Literature* 52, no. 1: 124–88.

McLeay, Michael, and Silvana Tenreyro. 2019. "Optimal Inflation and the Identification of the Phillips Curve." *NBER Macroeconomics Annual* 34:199–255.

Mertens, Thomas M., and John C. Williams. 2019. "Monetary Policy Frameworks and the Effective Lower Bound on Interest Rates." In *American Economic Association Papers and Proceedings*, vol. 109. Pittsburgh, Pa.: AEA Publications.

Moulton, Brent R. 2018. *The Measurement of Output, Prices, and Productivity: What's Changed since the Boskin Commission?* Washington: Brookings Institution. https://www.brookings.edu/research/the-measurement-of-output-prices-and-productivity/.

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

Orphanides, Athanasios, and John C. Williams. 2005. “Inflation Scares and Forecast-Based Monetary Policy.” *Review of Economic Dynamics* 8, no. 2: 498–527.

Peach, Richard, Robert Rich, and M. Henry Linder. 2013. “The Parts Are More than the Whole: Separating Goods and Services to Predict Core Inflation.” *Current Issues in Economics and Finance* 19, no. 7: 1–8.

Powell, Jerome H. 2019. “Challenges for Monetary Policy.” Speech given at the Economic Policy Symposium: Challenges for Monetary Policy, Federal Reserve Bank of Kansas City, Jackson Hole, Wyo., August 23.

Primiceri, Giorgio E. 2006. “Why Inflation Rose and Fell: Policy-Makers’ Beliefs and U. S. Postwar Stabilization Policy.” *Quarterly Journal of Economics* 121, no. 3: 867–901.

Primiceri, Giorgio E. 2019. Comment on “A Unified Approach to Measuring *u**,” by Richard K. Crump, Stefano Eusepi, Marc Giannoni, and Ayşegül Şahin. *Brookings Papers on Economic Activity*, Spring, 223–230.

Rognlie, Matthew. 2019. “Comment on Optimal Inflation and the Identification of the Phillips Curve.” *NBER Macroeconomics Annual* 34:267–79.

Rubbo, Elisa. 2020. “Networks, Phillips Curves and Monetary Policy.” Working Paper. https://scholar.harvard.edu/files/elisarubbo/files/rubbo_jmp.pdf.

Sbordone, Argia M. 2002. “Prices and Unit Labor Costs: A New Test of Price Stickiness.” *Journal of Monetary Economics* 49, no. 2: 265–92.

Sbordone, Argia M. 2007. “Globalization and Inflation Dynamics: The Impact of Increased Competition.” In *International Dimensions of Monetary Policy*, edited by Jordi Galí and Mark J. Gertler. Chicago: University of Chicago Press.

Smets, Frank, and Rafael Wouters. 2007. “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach.” *American Economic Review* 97, no. 3: 586–606.

Stansbury, Anna, and Lawrence H. Summers. 2020. “The Declining Worker Power Hypothesis: An Explanation for the Recent Evolution of the American Economy.” In the present volume of *Brookings Papers on Economic Activity*.

Stock, James H. 2011. Comment on “Inflation Dynamics and the Great Recession,” by Laurence Ball and Sandeep Mazumder, *Brookings Papers on Economic Activity*, Spring, 387–402.

Stock, James H., and M. W. Watson. 2007. “Why Has U.S. Inflation Become Harder to Forecast?” *Journal of Money, Credit and Banking* 39, no. s1: 3–33.

Stock, James H., and M. W. Watson. 2008. “Phillips Curve Inflation Forecasts.” Working Paper 14322. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w14322.

Stock, James H., and M. W. Watson. 2019. “Slack and Cyclically Sensitive Inflation.” Working Paper 25987. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25987.

Swanson, Eric T., and John C. Williams. 2014. “Measuring the Effect of the Zero Lower Bound on Medium- and Longer-Term Interest Rates.” *American Economic Review* 104, no. 10: 3154–85.

Tallman, Ellis W., and Saeed Zaman. 2017. “Forecasting Inflation: Phillips Curve Effects on Services Price Measures.” *International Journal of Forecasting* 33, no. 2: 442–57.

Taylor, John B. 1980. “Aggregate Dynamics and Staggered Contracts.” *Journal of Political Economy* 88, no. 1: 1–23.

Taylor, John B. 1993. “Discretion versus Policy Rules in Practice.” *Carnegie-Rochester Conference Series on Public Policy* 39:195–214.

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, N.J.: Princeton University Press.

Yellen, Janet. 2019. “Former Fed Chair Janet Yellen on Why the Answer to the Inflation Puzzle Matters.” Remark, Hutchins Center on Fiscal and Monetary Policy at Brookings, October 3. https://www.brookings.edu/research/former-fed-chair-janet-yellen-on-why-the-answer-to-the-inflation-puzzle-matters/.

Zhang, Bo, Joshua C. C. Chan, and Jamie L. Cross. Forthcoming. “Stochastic Volatility Models with ARMA Innovations: An Application to G7 Inflation Forecasts.” *International Journal of Forecasting*.

Comments and Discussion

COMMENT BY

OLIVIER BLANCHARD[1] This is a very nice paper. On the methodological front, it moves away from the traditional single equation estimation of the Phillips curve relation and shows the many useful ways in which one can go back and forth between vector autogression (VAR) representations and dynamic stochastic general equilibrium (DSGE) models to think about the relation between inflation and unemployment. On the conclusions front, the paper largely confirms current wisdom, which is reassuring. It reaches three main conclusions: The relation between inflation and unemployment reflects primarily a causal effect of unemployment on inflation. The effect of unemployment on price inflation has become weaker. The effect of unemployment on wage inflation has also become weaker but less so.

In my comments, I shall focus mostly on one aspect of their results, the adjustment of prices and wages to movements in unemployment. I shall argue that we should focus more on the movements in markups—the ratio of prices to nominal unit labor costs—and think about their policy implications.

First, however, I shall briefly comment on two aspects of their methodology.

The first is, rightly, a focus on potential reverse causality from inflation to unemployment, leading to a correlation between the error term and unemployment. A recent paper by McLeay and Tenreyro (2019) has argued that a change in monetary policy could indeed be the explanation behind the change in the Phillips curve relation. To explore the issue, the authors use the VAR equivalent of an instrumental variable single equation approach,

1. Thanks to Mark Bils for discussions and to Chris Erceg for simulations.

by looking at the co-movements in inflation and unemployment conditional on a demand shock. They take as demand shocks movements in the excess bond premium, an unusual choice and one which strikes me as somewhat unconvincing. The factors behind the movements in the premium, namely, variations in market risk aversion, may well have both strong supply and demand effects. I would have preferred the use of a more conventional instrument, such as movements in cyclically adjusted fiscal balances or Romer-Romer fiscal series, which seem more likely to affect demand than supply in the short run. The authors conclude that reverse causality does not appear to play a major role in explaining the change in the relation between unemployment and inflation. I suspect the use of a more standard instrument would lead to the same conclusion.

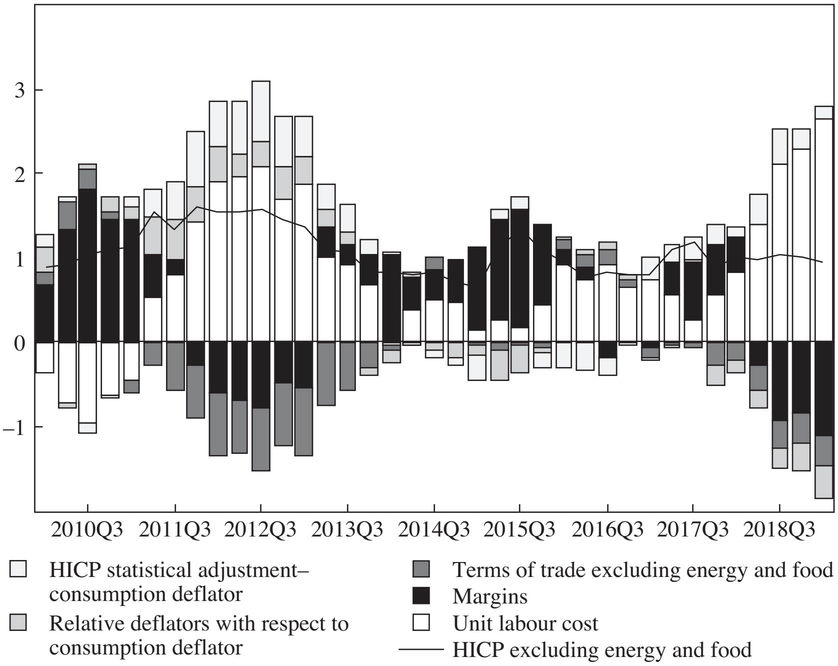
The second is a master class in how to go from reduced form VARs to structural VARs to estimated DSGEs and back. The idea of using VARs to derive impulse price responses conditional on similar impulse responses for unemployment across two samples is clever—if probably subject to a Lucas critique. Comparing implied VAR representations from DSGE models to actual VARs under different assumptions about parameters to get a sense of how different sets of assumptions fit various aspects of the data is also clever and well done. I learned a lot from the paper.

Turning now to the conclusions. The various approaches used in the paper all suggest that the price response to unemployment has significantly decreased over time. The evidence on the response of wages to unemployment is less conclusive. Their VAR approach suggests that the wage response has also decreased, but less than that of prices. The DSGE approach gives a stronger, more statistically significant, answer, namely, that the wage response has decreased by less than the price response.

I want to focus in my comments on the behavior of markups. To come to a conclusion about the behavior of markups, however, one needs to know the response of nominal unit costs—and thus not only the response of wages but also the response of productivity—to unemployment. In private conversation, the authors have kindly given me the response of productivity, which looks largely similar in both samples. I conclude that the authors provide some evidence of countercyclical markups: when unemployment increases, wages decrease, but markups increase, more so than they used to, leading to only a muted response of prices.

I shall take this as license to think about the cyclical behavior of markups and its policy implications. I do so because there is substantial evidence that the behavior of markups is more complex and more important than we used to believe. In the Dixit-Stiglitz model of monopolistic

Figure 1. Breakdown of the Year-on-Year Change in Core Inflation in the Euro Area



Source: Diev, Kalantzis, and Lalliard (2019).

competition—still a standard building block of most DSGE models—the elasticity of substitution between goods is constant, and so is the markup. Reality suggests a more complex picture.

Let me go through two additional pieces of evidence.

The first is the evolution of markups in the euro zone. Figure 1, taken from Diev, Kalantzis, and Lalliard (2019), shows the evolution of core inflation since 2010 in the euro zone and its decomposition between terms of trade effects, nominal unit labor costs, and markups. Concentrating on the end of the sample, two aspects are striking. The first is the increase in unit labor costs inflation since 2015, reflecting (behind the scene) the pressure of lower unemployment on wage inflation. The second is how this increase in cost inflation has not translated in price inflation: as costs have increased, markups have decreased nearly in unison. In other words, markups have been strongly countercyclical, decreasing as output increased and unemployment decreased.

The second piece of evidence builds on the work of Bils, Klenow, and Malin (2018), who argued that markups were strongly countercyclical. Following the approach of their paper, I construct series for markups, real output, and nominal unit labor costs in the US nonfarm business sector. Markups are constructed as the ratio of the price deflator to the nominal unit labor costs, thus as the inverse of the labor share. The results of a regression of log markups on log real output and log nominal unit costs are given for two subsamples, 1980–1999 and 2000–2019. All three series are hp-filtered.

$$\begin{array}{ll} 1980\text{–}1999 & \mu = -0.10y - 0.08ulc + \epsilon \\ & R^2 = 0.09, \sigma(\mu) = 1.2 \text{ percent} \\ 2000\text{–}2019 & \mu = -0.23^{**}y - 0.60^{**}\,ulc + \epsilon \\ & R^2 = 0.69, \sigma(\mu) = 1.9 \text{ percent} \end{array}$$

**significant at the 95 percent confidence level

In both samples, the coefficients on output and nominal unit labor costs are negative. Markups indeed appear to be countercyclical, and an increase in nominal costs leads to a decrease in markups, suggesting real or nominal rigidities. The results are, however, significantly different across the two samples. In the earlier sample, the coefficients are small and insignificant. In the later sample, both coefficients are large and strongly significant. Also, the variation in markups, measured by the standard deviation, is substantially higher in the later sample.

This difference across the two samples is robust to other ways of detrending, such as first differencing, and other ways of measuring marginal cost, not using wages, which Bils, Klenow, and Malin (2018) insisted might not represent the marginal cost of labor but rather the price of materials. It suggests that, for various reasons, prices have become more inertial, and markups more variable than they used to be. I shall not explore what the reasons might be, but take these two facts as stylized facts and draw potential policy implications. Let me state the bottom line: both increased price rigidity and larger markup shocks strengthen the case for wage inflation targeting over price inflation targeting.

Consider the simplistic nominal wage and price equations:

$$w = ap + (1 - a)Ep - cu + \eta$$

$$p = bw + (1 - b)Ew + \epsilon.$$

The coefficients a and b proxy for the stickiness of wages and prices. The lower a or b, the more predetermined prices or wages. Assume ϵ and η are white noise and capture distortions, so that the optimal level of unemployment is invariant to any of the two shocks and, given the white noise assumptions, equal to zero.

Consider now the effect of a markup shock, $\epsilon > 0$. Given the white noise assumption, Ep and Ew are both equal to zero, so the two equations become:

$$w = ap - cu$$

$$p = bw + \epsilon.$$

Now consider the effects of the shock under two alternative monetary regimes, price inflation targeting, $p = 0$, and wage inflation targeting, $w = 0$.

Under price inflation targeting, unemployment is given by:

$$u = \frac{1}{bc}\epsilon.$$

Under wage inflation targeting, unemployment is given by:

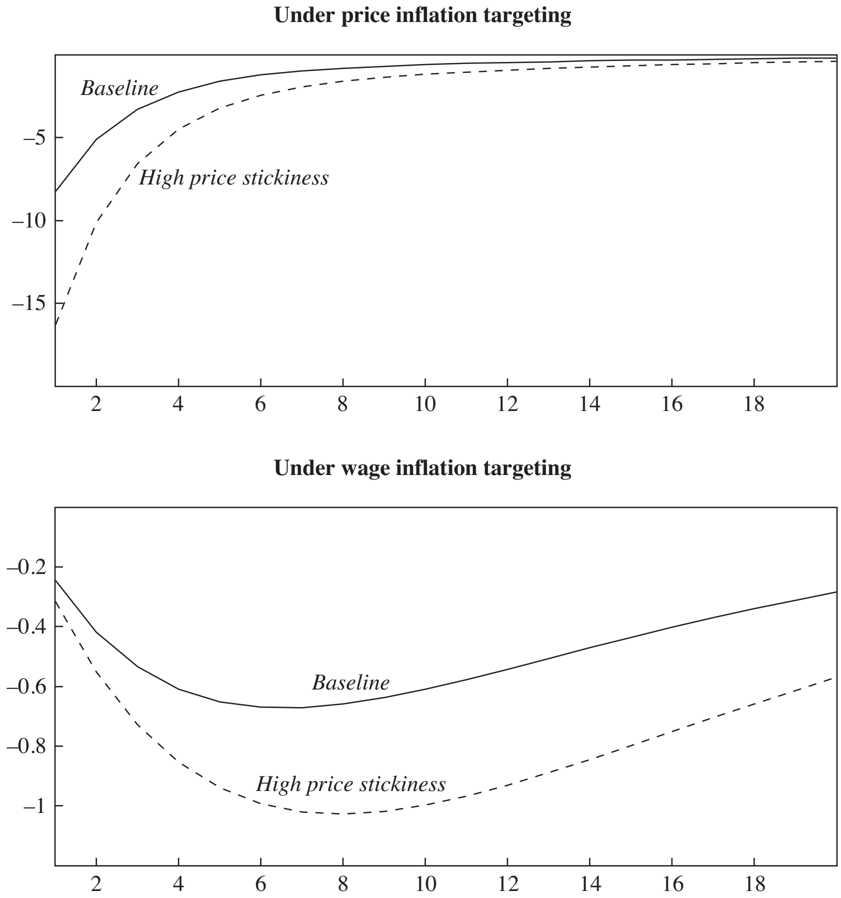
$$u = \frac{a}{c}\epsilon.$$

Given that $a \leq 1$, the implication is that, in response to markup shocks, wage inflation targeting always dominates price inflation targeting. And that the higher the price rigidity (the lower b) and the higher the variance of markup shocks, the more wage inflation targeting dominates.

The intuition is straightforward. In response to a positive markup shock, maintaining a stable price level requires a decrease in the nominal wage, and this decrease requires in turn an increase in unemployment. The less the price responds to the wage, the more the wage has to decrease and the more the unemployment rate has to increase. In contrast, maintaining a stable wage level, and thus allowing the price to increase with the markup, requires a smaller increase in unemployment.

The same conclusions follow from simulations of a larger model. The simulations below—courtesy of Chris Erceg, with whom I am doing joint work on this issue—show the effects of a markup shock in a midsize New Keynesian model with nominal wage and nominal price rigidities. The simulations show the effect of an AR(1) markup shock, with AR

Figure 2. Output Response to a Markup Shock



Source: Author's calculations.

coefficient 0.9, under either strict price inflation targeting or strict wage inflation targeting, and in each case, with either baseline or high price rigidity (captured by the Calvo coefficient on the expected length of price setting).

Figure 2 yields two conclusions. First, under the baseline, the output cost is much lower under wage inflation than price inflation targeting, –0.2 percent on impact and –0.6 percent at the trough under wage inflation targeting, versus –8.0 percent on impact under price inflation targeting. Second, the output cost under price inflation targeting is much larger when there is high price rigidity, −16 percent versus −8 percent under the baseline.

Let me finally turn from the effects of markup shocks to the effects of demand shocks. Extend the simplistic model above to add an aggregate demand equation and an interest rate rule:

$$w = ap + (1-a)Ep - cu + \eta$$

$$p = bw + (1-b)Ew + \epsilon$$

$$u = r + \epsilon_u$$

$$r = \alpha\big(dp + (1-d)w\big).$$

Unemployment is an increasing function of the policy rate and a demand shock, ϵ_u. The policy rate responds to both the price and the wage. One can think of price inflation targeting as $d = 1$ and wage inflation targeting as $d = 0$. Consider an unexpected demand shock, ϵ_u. Solving for unemployment under price inflation targeting gives:

$$u = \frac{1-ab}{1-ab+cb\alpha}\epsilon_u.$$

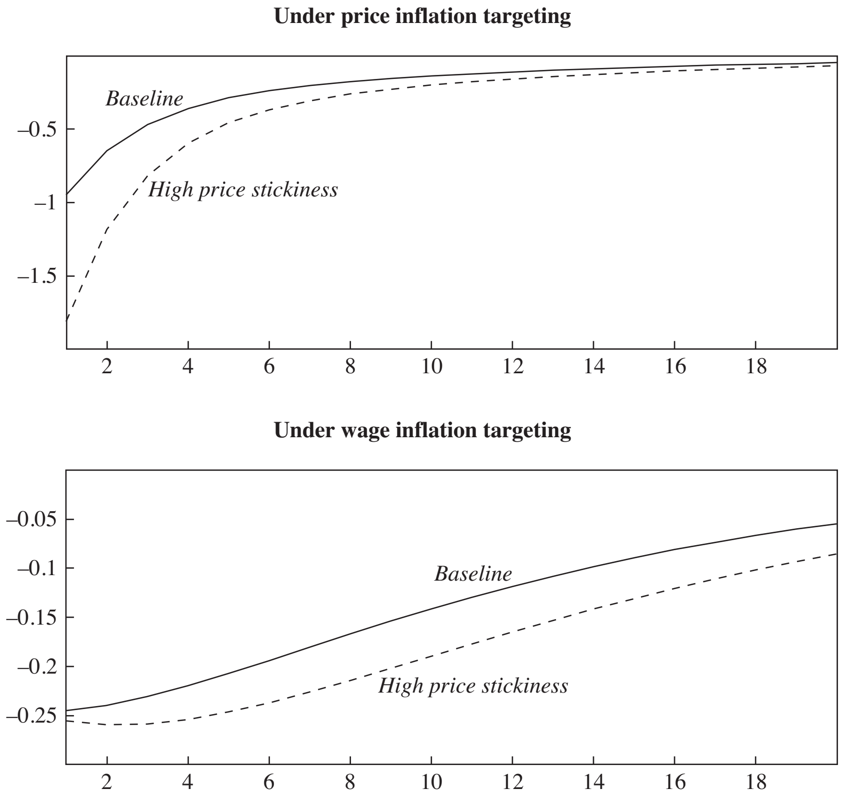
Solving for unemployment under wage inflation targeting gives instead:

$$u = \frac{1-ab}{1-ab+c\alpha}\epsilon_u.$$

So long as $b < 1$, the effect of the demand shock on unemployment is smaller under wage than price inflation. The intuition is again straightforward. If prices respond little to wages and thus respond little to unemployment, and if the interest rate responds to prices, the monetary response will be small, and demand shocks will have a large effect on unemployment. This is particularly clear if one considers the case where $b = 0$. Under wage targeting, $u = (1/(1 + c\alpha))\epsilon_u$, but under price targeting, $u = \epsilon_u$; in that case, as prices do not react, monetary policy does not react either, and the demand shock has a full effect on unemployment.

Again, it is useful to look at the question in a larger model. Based on the same model as above, figure 3 shows the effect of an AR(1) demand shock, with AR coefficient 0.9, under price inflation targeting and under wage inflation targeting and, in each case, under baseline and high price rigidity. It yields two conclusions. The effect of the shock on output is substantially

Figure 3. Output Response to a Demand Shock



Source: Author's calculations.

smaller under wage targeting. And the higher the price rigidity, the larger the comparative advantage of wage over price inflation targeting.

The implications of nominal price rigidity for the effects of demand shocks are less robust than for the effects of markup shocks: in principle, a smaller response of prices to unemployment can be offset by a more aggressive interest rate rule (a larger α in the model above). But if the signal from prices is noisy because of movements in markups, a more aggressive rule will lead to too strong a reaction to a markup shock, an undesirable outcome.

These comments are not the place to go further in exploring wage inflation targeting. But I see the conclusions of the paper, as well as a body of other evidence, as suggesting that the nature of the Phillips curve has changed, that the behavior of prices given unit labor costs has changed, and that this is ground enough to explore wage inflation targeting.

REFERENCES FOR THE BLANCHARD COMMENT

Bils, Mark, Peter Klenow, and Benjamin Malin. 2018. "Resurrecting the Role of the Product Market Wedge in Recessions." *American Economic Review* 108, nos. 4-5: 1118–46.

Diev, Pavel, Yannick Kalantzis, and Antoine Lalliard. 2019. "Why Have Strong Wage Dynamics Not Pushed Up Inflation in the Euro Area?" Bulletin de la Banque de France 225/6, September-October. https://publications.banque-france.fr/en/why-have-strong-wage-dynamics-not-pushed-inflation-euro-area.

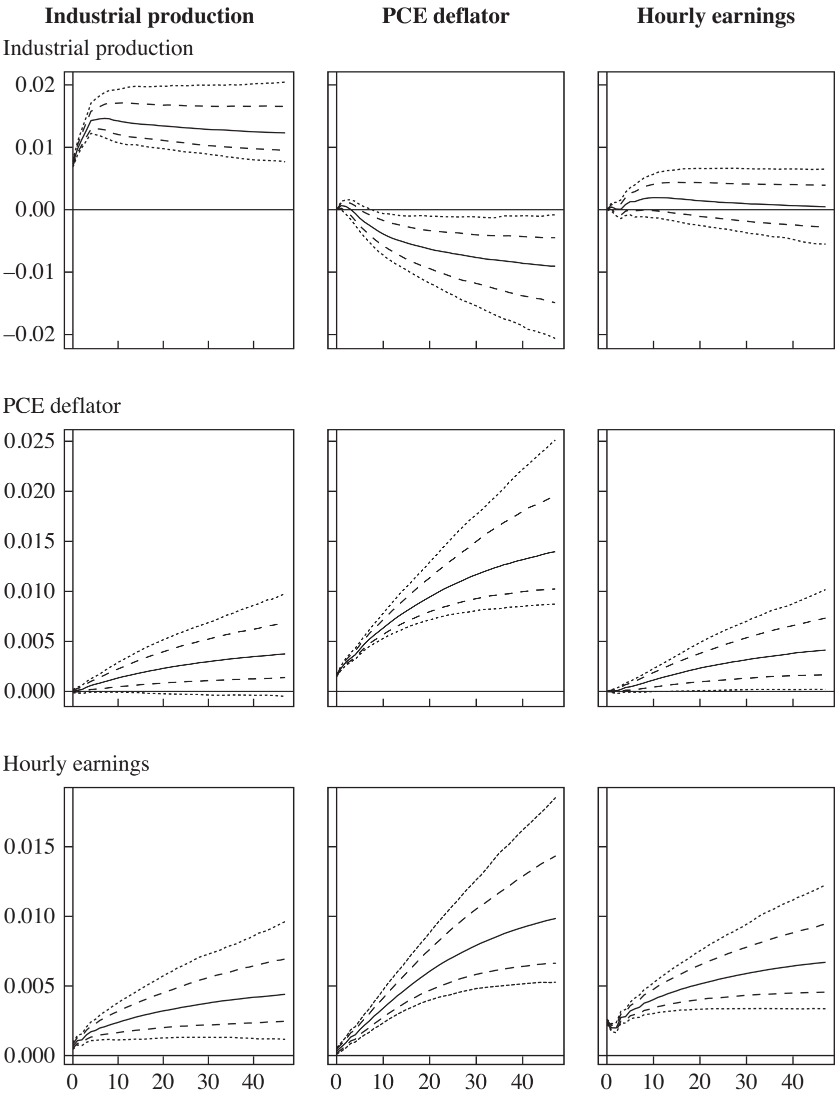
McLeay, Michael, and Silvana Tenreyro. 2019. "Optimal Inflation and the Identification of the Phillips Curve." *NBER Macroeconomics Annual* 34:199–255.

COMMENT BY

CHRISTOPHER A. SIMS The main fact the paper documents is that the relation between a one-dimensional measure of real activity and the rate of wage or price inflation has weakened between pre-1990 and post-1990 periods in the United States. This is a robust result, emerging with a variety of measures of real activity and a variety of time series modeling approaches. But despite using tools capable of unraveling a richer story, the paper maintains a one-dimensional conceptual framework, with a single driving force for a "business cycle" that moves all variables in a repeating pattern. The paper acknowledges that this does not account for all variation in inflation but suggests that longer-run variation in inflation can be set aside when focusing on its "cyclical" variation. This leads the paper into a discussion of why inflation has become less variable over a long span of time, while making no effort to account for the fact that inflation's level has come steadily down over this span of time and has recently been persistently below Federal Reserve targets.

Most variation in inflation is not movement along a Phillips curve. Since the paper does not display impulse responses to disturbances other than its composite unemployment-shifting shock, I have estimated a simple three-variable vector autoregression (VAR), with monthly data on industrial production, the personal consumption expenditures (PCE) deflator, and average hourly earnings of production and nonsupervisory employees. The impulse responses, pre- and post-1990, are displayed in figures 1 and 2.

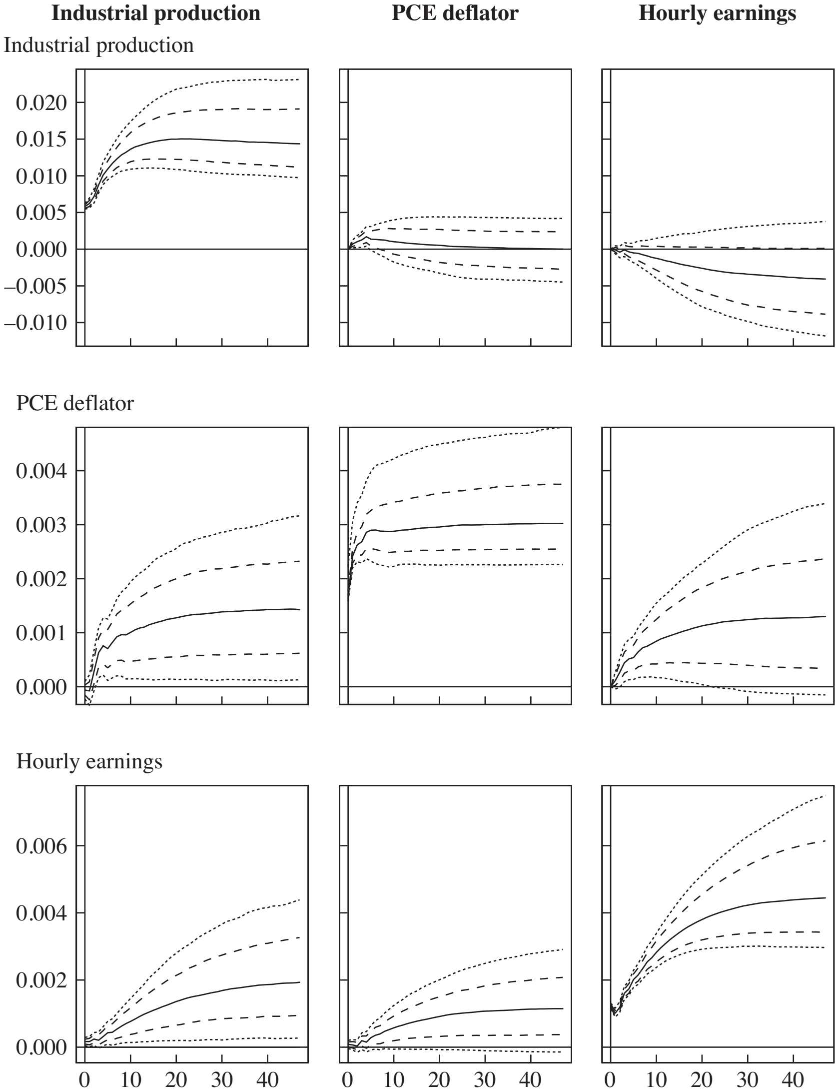
Figure 1. Pre-1990 Impulse Responses



Source: Author's calculations.

Notes: Impulse responses from a VAR estimated with monthly data for industrial production, PCE deflator, and average hourly earnings for January 1965 through December 1989. The impulse responses have been orthogonalized by Cholesky decomposition, with industrial production first in the ordering. Estimation used a Minnesota prior. The two error bands are 68 percent and 90 percent bands.

Figure 2. Post-1990 Impulse Responses



Source: Author's calculations.

Notes: Impulse responses from a VAR estimated with monthly data for industrial production, PCE deflator, and average hourly earnings for January 1990 through May 2016. The impulse responses have been orthogonalized by Cholesky decomposition, with industrial production first in the ordering. Estimation used a Minnesota prior. The two error bands are 68 percent and 90 percent bands.

The first column of each of these figures captures the main message of the paper: there is a disturbance that moves output, prices, and wages in the same direction. Scanning across the first row, we can see that this first-column shock accounts for most of the variation in output. (All plots in each row have the same scale, so that visually smaller responses account for less variation in the row variable.) The size of the first-column disturbance to industrial production is about the same in both plots, with the median estimate of long-run response of industrial production (at the end of the five years shown) around 1.4 percent in both periods. The scale of the price and wage rows of the plots is quite different in the two periods. The median estimate of the long-run response of the PCE deflator to this first-column shock is around 0.12 percent in the post-1990 period and around 0.40 percent in the earlier period. For the wage response the corresponding responses are 0.20 percent and 0.5 percent. While these differences are in the same direction as found in the paper, it is worth noting that the 68 percent error bands overlap.

What is omitted from the paper is the message of the bottom two rows of the plot arrays. For both prices and wages, in both time periods, most of the variation is being generated by disturbances that do not move output in the same direction as wages and prices. In the pre-1990 case, the second column, which accounts for most of the variance in prices and wages, is a shock that moves output down and both prices and wages up. This probably reflects the oil shock stagflation of the 1970s. In the post-1990 case the corresponding shock is in the third column. It also moves output down and inflation and wages up, though the output decline is only marginally statistically significant. The second column post-1990 and the third column pre-1990 are similar in showing almost no response of industrial output and substantial price and wage responses, though the relative importance of these shocks for explaining price inflation is greater after 1990.

The responses in the second two rows are all at a smaller scale after 1990, but the relative importance for explaining wages and prices of the first column shock, which behaves like a movement along a Phillips curve, is about the same in both periods. These responses do fit the story that the slope of the Phillips curve—the response of wages and prices to the level of business activity—has declined. They do not fit the additional claim in the paper that this is the main source of the decline in variability of inflation.

The United States, Japan, and the euro area all moved steadily toward zero interest rates and below-target inflation after 1990. Benhabib, Schmitt-Grohé, and Uribe (2001) showed us that, once the zero bound on interest rates is recognized, the standard modeling of inflation dynamics with a Taylor rule policy leads to a drift toward near-zero interest rates and low or negative inflation. Their argument depends on there not being a reliable fiscal expansion response to the occurrence of the low interest rates, but in light of recent experience this seems realistic.

The Benhabib, Schmitt-Grohé, and Uribe (2001) model can explain persistent low inflation, persistent low interest rates, inability of the Federal Reserve to affect inflation, and insensitivity of inflation to real disturbances. It does not explain the high levels of real government debt in the United States or the low real interest rates. We don't have a model that pulls all these facts together. But Benhabib, Schmitt-Grohé, and Uribe (2001) seem to offer a better starting point for understanding these facts than a narrow focus on the Phillips curve.

I also have one narrower criticism of the paper. I'm not convinced there is any substantial difference between wage and price inflation in the changes between the pre-1990 and post-1990 periods. Both in figure 3 and in figure 4 in the paper, the sizes of the responses of wage and price inflation are quite similar in the post-1990 period. In figure 2 in the paper the responses of wage inflation seem weaker, not stronger, than those of prices. So the claim that the wage Phillips curve is still important after 1990 does not seem supported by the reduced form statistical analysis. It depends on the results in section IV, which invokes the theoretical structure of the New York Federal Reserve model, and in particular the notion of separate wage and price Phillips curves with mutually uncorrelated shocks and driven by distinct real variables. The idea that workers have pricing power and control the quantity of their labor to maintain a wage markup, implicit in this specification of a wage Phillips curve, is in my view at best a modeling convenience. So this aspect of the paper's conclusions might need several grains of salt.

REFERENCE FOR THE SIMS COMMENT

Benhabib, Jess, Stephanie Schmitt-Grohé, and Martín Uribe. 2001. "The Perils of Taylor Rules." *Journal of Economic Theory* 96, no. 1–2: 40–69.

GENERAL DISCUSSION Rick Mishkin began by observing that the authors' conclusions about the slope of the Phillips curve are not consistent with evidence from analyses of Phillips curves using state- and city-level data in, for example, papers by McLeay and Tenreyro and by Hooper, Mishkin, and Sufi.[1] Importantly, disaggregated data allow these researchers to control for endogenous monetary policy because states and cities within a monetary union experience the same monetary policy setting. These analyses find that the Phillips curve slope did not change significantly after 1990, in contrast to the current paper's conclusions. Mishkin asked whether the authors predict a difference between the disaggregated and aggregate data that would help reconcile this micro evidence with the authors' finding that the Phillips curve slope declined after 1990.

Giovanni Ricco observed that the estimation of trend inflation and unemployment is a critical step in fitting a Phillips curve, which is primarily concerned with the relationship between the cyclical component of inflation and the cyclical component of slack. He commented that the vector autoregressions (VARs) used in the authors' main exercise would implicitly fit a deterministic trend to the data, potentially leading to misestimation of the cyclical components of inflation and unemployment. He asked whether the authors had examined whether this feature affects their results.

Robert Gordon made two comments. First, he reemphasized Sims's reference to supply shocks as crucial to understanding inflation before 1990. He noted, for example, that most of the variation in inflation between 1973 and 1985 can be explained by unfavorable supply shocks that occurred during the 1975 and 1980 oil crisis episodes. Similarly, the very sharp decline of inflation between 1981 and 1985 was due not just to high unemployment but also to the dollar's appreciation during that period. Gordon said controlling for these supply disturbances is important for understanding the shape of the Phillips curve and observed that they are not explicitly considered in the paper.

Second, Gordon observed that two broad hypotheses have been offered to explain why inflation has been subdued in the last ten years despite a dramatic decline in unemployment. One is that the Phillips curve has flattened, as the authors argue. The other is that the rate of unemployment consistent with stable inflation (NAIRU) has declined. The NAIRU hypothesis has been discussed extensively in the context of the low inflation of the 1990s, but Gordon observed that some evidence suggests the NAIRU has declined even further in the last decade. Globalization, the decline in worker bargaining power, the decline in computer prices, and the increased

1. Michael McLeay and Silvana Tenreyro, "Optimal Inflation and the Identification of the Phillips Curve," *NBER Macroeconomics Annual* 34 (2019): 199–255; Peter Hooper, Frederic S. Mishkin, and Amir Sufi, "Prospects for Inflation in a High Pressure Economy: Is the Phillips Curve Dead or Is It Just Hibernating?," *Research in Economics* 74, no. 1 (2020): 26–62.

importance of computer and IT-related goods are all examples of factors that could have reduced the NAIRU and subdued inflation.

However, he observed that a decline in the NAIRU cannot account for why inflation declined very modestly after the sharp increase in unemployment after the 2008 financial crisis or during the period of weak activity around 2015. Gordon noted that in a contest between the NAIRU and the flat Phillips curve hypotheses, he would likely fall in favor of the flat Phillips curve explanation, consistent with the authors' conclusions.

Giorgio Primiceri responded to comments from Ricco by noting that the authors' VAR approach allows them to measure the inflationary response specifically to an identified cyclical shock. More specifically, the authors verify that estimated trend variables like the NAIRU do not respond to the business cycle shock used in the VAR exercise, suggesting that the identifying disturbance is in fact a cyclical one.

Primiceri acknowledged comments from Sims that the evidence that the business cycle's correlation with wage inflation has declined less than its correlation with price inflation is somewhat tenuous and noted that this finding depends partly on the measure of wage inflation used. For example, the authors find that the correlation between unemployment and wage inflation measured by the employment cost index remains relatively strong before and after 1990; in contrast, the correlation between unemployment and average hourly earnings declined about in line with the unemployment-price inflation correlation after 1990.

Primiceri acknowledged that the business cycle shock does not capture all the relevant historical business cycle variation and that it may not capture important movements in lower-frequency variables that affect inflation. However, the authors' focus in this paper was to examine why inflation does not move over the business cycle as much as it did in the past; the shock was chosen to help address this question.

In addition, while Primiceri acknowledged the excess bond premium shock used in the VAR exercise is a mix of a demand and supply shock, as Sims and Blanchard both suggested, he also noted that were one to replace it with a more traditional demand shock, like an unemployment shock, the main results would hold. The authors' interpretation of the response of inflation to the shock remains valid if the shock contains a prevalent demand component, which the literature surrounding the excess bond premium suggests it does.

Andrea Tambalotti responded to comments from Blanchard, acknowledging that the dynamics of markups are very important for explaining inflation's behavior and perhaps more so in the most recent period. He also noted that targeting wage inflation, which appears to have a stronger correlation with the business cycle, would be an interesting policy in the environment of flatter aggregate supply. However, he referenced work by himself, Justiniano, and Primiceri that examines the quantitative losses associated with stabilizing wage inflation and stabilizing price inflation and finds that the two strategies perform similarly.[2] That result might hint that the wedges between wages and prices are not terribly important for explaining the disconnect between inflation and the business cycle.

Marco Del Negro first addressed comments by both discussants that the business cycle shock is somewhat simplistic. He stressed that the authors test the response of inflation to multiple shocks throughout the paper. In particular, the exercise using the estimated dynamic stochastic general equilibrium (DSGE) model features a large suite of shocks to many variables in the model, and the results from that exercise are very much consistent with the results from the VAR exercise with one business cycle shock.

Del Negro also replied to comments from Mishkin, referencing work by Hazell, Herreno, Nakamura, and Steinsson, who point out that regressions of state- or city-level inflation data on unemployment may be confounded by the fact that different locations feature different unemployment and adjustment dynamics.[3] Controlling for these differences across locations could be possible with data on, for example, location-specific inflation expectations or structural unemployment, but these data are not readily available.

He also responded to Blanchard's comment that markups have become much more countercyclical in the last twenty years than they were prior. Consistent with this observation, Del Negro affirmed that the authors' DSGE model estimated on post-1990 data suggests that real marginal costs (the inverse of aggregate markups) move quite procyclically. In addition, consistent with Blanchard's comment about targeting wage inflation, he noted that systematic policy in the DSGE model is very effective at moving marginal costs precisely because of their stronger relation with the business cycle.

2. Alejandro Justiniano, Giorgio E. Primiceri, and Andrea Tambalotti, "Is There a Trade-Off between Inflation and Output Stabilization?," *American Economic Journal: Macroeconomics* 5, no. 2 (2013) 1–31.

3. Jonathon Hazell, Juan Herreno, Emi Nakamura, and Jón Steinsson, "The Slope of the Phillips Curve: Evidence from US States," Working Paper (2020).