BROOKINGS

www.brookings.edu

$9.95
ISBN 978-0-8157-3815-2

Brookings Papers ON ECONOMIC ACTIVITY

BPEA
SPRING
2019

BROOKINGS

Janice Eberly and James H. Stock, Editors

Brookings Papers
ON ECONOMIC ACTIVITY

SPRING 2019

RACHEL and SUMMERS
on Secular Stagnation in the Industrialized World

CHEN, CHEN, HSIEH, and SONG
on a Forensic Examination of China's National Accounts

CRUMP, EUSEPI, GIANNONI, and ŞAHIN
on a Unified Approach to Measuring u^*

ROMER and ROMER
on Fiscal Space and the Aftermath of Financial Crises: How It Matters and Why

AARONSON, DALY, WASCHER, and WILCOX
on Okun Revisited: Who Benefits Most from a Strong Economy?

METCALF
on the Economics of a Carbon Tax for the United States

Brookings Papers

ON ECONOMIC ACTIVITY

SPRING 2019

JANICE EBERLY
JAMES H. STOCK
Editors

BROOKINGS INSTITUTION PRESS
Washington, D.C.

ISSN 0007-2303
ISBN: Print 978-0-8157-3815-2; ebook: 978-0-8157-3816-9

Brookings Papers
ON ECONOMIC ACTIVITY

SPRING 2019

This journal is printed in black and white, but is available online in color. Often, the numbered graphical figures in a paper may contain colored elements (lines, dots, bars, and areas), to which the paper's text refers. To see these colored elements, please consult the online versions of the papers, at https://www.brookings.edu/project/brookings-papers-on-economic-activity/.

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 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 Financial Management, Inc.; 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*—along with appendix materials, data, and programs used to generate results—are made freely available for download at www.brookings.edu/past-bpea-editions. 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 SEVENTH CONFERENCE

Stephanie R. Aaronson *Brookings Institution*
Katharine Abraham *University of Maryland*
George Akerlof *Georgetown University*
Martin Neil Baily *Brookings Institution*
Susanto Basu *Boston College*
Joe Beaulieu *Brevan Howard*
Ben S. Bernanke *Brookings Institution*
Olivier Blanchard *Peterson Institute for International Economics*
William Brainard *Yale University*
Steven Braun *Council of Economic Advisers*
Jesse Bricker *Federal Reserve Board of Governors*
Ralph Bryant *Brookings Institution*
Elaine Buckberg *General Motors*
Marcus Casey *Brookings Institution*
David Cashin *Federal Reserve Board of Governors*
Wei Chen *Chinese University of Hong Kong*
Xilu Chen *Chinese University of Hong Kong*
Gerald Cohen *Haver Analytics*
Charles Collyns *International Monetary Fund*
Richard Cooper *Harvard University*
Richard K. Crump *Federal Reserve Bank of New York*
Mary C. Daly *Federal Reserve Bank of San Francisco*
Troy Davig *Rokos Capital Management*
Steven J. Davis *University of Chicago*
Jane Dokko *Federal Reserve Bank of Chicago*
David Dollar *Brookings Institution*
Janice Eberly *Northwestern University*
Wendy Edelberg *Congressional Budget Office*
Gauti B. Eggertsson *Brown University*
Douglas Elmendorf *Harvard Kennedy School*
Stefano Eusepi *University of Texas at Austin*
Stanley Fischer *BlackRock*
Martin Fleming *IBM*
Renee Fry-McKibbin *Australian National University*
Jason Furman *Harvard Kennedy School*
William Gale *Brookings Institution*
Ted Gayer *Brookings Institution*
Domenico Giannone *Federal Reserve Bank of New York*
Marc Giannoni *Federal Reserve Bank of Dallas*
Robert J. Gordon *Northwestern University*
Tracy Gordon *Urban Institute*
Josh Gotbaum *Brookings Institution*
Lawrence Goulder *Stanford University*
Robert E. Hall *Stanford University*
John Haltiwanger *University of Maryland*
Harry Holzer *Georgetown University*
Julie L. Hotchkiss *Federal Reserve Bank of Atlanta*
Caroline Hoxby *Stanford University*
Chang-Tai Hsieh *University of Chicago*
Bo Yeon Jang *Federal Reserve Board of Governors*
Michael Kiley *Federal Reserve Board of Governors*
Jeffrey Kling *Congressional Budget Office*
Donald Kohn *Brookings Institution*
Arvind Krishnamurthy *Stanford University*

Adam Looney *Brookings Institution*
Christos Makridis *Massachusetts Institute of Technology–Sloan School of Management*
N. Gregory Mankiw *Harvard University*
Donald Marron *Urban Institute*
Mark Mazur *Urban-Brookings Tax Policy Center*
Robert McClelland *Urban Institute*
Warwick McKibbin *Australian National University*
Karel Mertens *Federal Reserve Bank of Dallas*
Gilbert E. Metcalf *Tufts University*
Laurence Meyer *Monetary Policy Analytics*
Frederic Mishkin *Columbia University*
Adele C. Morris *Brookings Institution*
Edward Nelson *Federal Reserve Board of Governors*
Ryan Nunn *Brookings Institution*
Tom Orlik *Bloomberg*
George Perry *Brookings Institution*
Jonathan Pingle *BlackRock*
Eswar Prasad *Brookings Institution*
Giorgio E. Primiceri *Northwestern University*
Łukasz Rachel *Bank of England*
Valerie Ramey *University of California, San Diego*
David Ratner *Federal Reserve Board of Governors*
John Roberts *Federal Reserve Board of Governors*
Christina D. Romer *University of California, Berkeley*
David H. Romer *University of California, Berkeley*
Ayşegül Şahin *University of Texas at Austin*
Natasha Sarin *University of Pennsylvania*
Jay Shambaugh *Brookings Institution*
Louise Sheiner *Brookings Institution*
Tara Sinclair *George Washington University*
Zheng Song *Chinese University of Hong Kong*
Mark Steinmeyer *Smith Richardson Foundation*
James H. Stock *Harvard University*
Lawrence H. Summers *Harvard University*
Phillip Swagel *University of Maryland*
Marzie Taheri Sanjani *International Monetary Fund*
Robert Tetlow *Federal Reserve Board of Governors*
Alan Viard *American Enterprise Institute*
Ivan Vidangos *Federal Reserve Board of Governors*
William L. Wascher *Federal Reserve Board of Governors*
David Wessel *Brookings Institution*
David W. Wilcox *Federal Reserve Board of Governors (Retired)*
Justin Wolfers *University of Michigan*
Jonathan Wright *Johns Hopkins University*
Wei Xiong *Princeton University*
Janet Yellen *Brookings Institution*
Jeromin Zettelmeyer *Peterson Institute for International Economics*

Francisca P. Alba *Brookings Institution*
Alfred F. Imhoff *Brookings Institution*
Mary E. King *Brookings Institution*

ŁUKASZ RACHEL
Bank of England

LAWRENCE H. SUMMERS
Harvard University

On Secular Stagnation in the Industrialized World

ABSTRACT We argue that the economy of the industrialized world, taken as a whole, is currently—and for the foreseeable future will remain—highly prone to secular stagnation. But for extraordinary fiscal policies, real interest rates would have fallen much more and be far below their current slightly negative level, current and prospective inflation would be further short of the 2 percent target levels, and past and future economic recoveries would be even more sluggish. We start by arguing that, contrary to current practice, neutral real interest rates are best estimated for the bloc of all industrial economies, given capital mobility between them and the relatively limited fluctuations in their aggregated current account. We show, using standard econometric procedures and looking at direct market indicators of prospective real rates, that neutral real interest rates have declined by at least 300 basis points over the last generation. We argue that these secular movements are in larger part a reflection of changes in saving and investment propensities rather than the safety and liquidity properties of Treasury instruments. We highlight the observation that, ceteris paribus, levels of government debt, the extent of pay-as-you-go old-age pensions, and the insurance value of government health

Conflict of Interest Disclosure: Łukasz Rachel is a senior economist at the Bank of England. An earlier version of this paper was one of the chapters in his thesis for the degree of doctor of philosophy at the London School of Economics and Political Science. Lawrence Summers is the Charles W. Eliot 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 the Bank of England, the London School of Economics and Political Science, or Harvard University.

care programs have all operated to raise neutral real rates. Using estimates drawn from the literature—as well as two general equilibrium models emphasizing, respectively, life-cycle heterogeneity and individual uncertainty—we suggest that the "private sector neutral real rate" may have declined by as much as 700 basis points since the 1970s. The extent of the substantial shifts in private saving and investment propensities over time has been obscured by the impact of this decline in real rates. Our diagnosis necessitates radical revisions in the conventional wisdom about monetary policy frameworks, the role of fiscal policy in macroeconomic stabilization, and the appropriate level of budget deficits, as well as social insurance and regulatory policies. To that end, much more of creative economic research is required on the causes, consequences, and policy implications of the pervasive private sector excess saving problem.

Long after the 2008 financial crisis, real interest rates in the economies of the industrialized world remain very low by recent historical standards, central banks' balance sheets are inflated, government debt and deficit levels are high, and yet nominal GDP growth remains too low for the achievement of 2 percent inflation targets. This has led to a revival of interest in the secular stagnation hypothesis, according to which a chronic tendency of private investment to be insufficient to absorb private saving leads, in the absence of extraordinary policies, to extremely low interest rates, inflation that is lower than desirable, and sluggish economic growth.

Much of the discussion has focused on movements in what has come to be called "*R*-star" (R^*)—Knut Wicksell's (1898) neutral or natural interest rate, at which investment fully absorbs saving at full employment. Estimating the level and change in the neutral real interest rate has become a cottage industry, and the neutral real interest rates have come to play a prominent role in policy discussions.

Policymakers have taken notice of significant movements in the apparent neutral real rate. Federal Reserve chairman Jerome Powell's recent remark that the nominal federal funds rate—at the time, set at between 2 and 2.25 percent—was "just below the broad range of estimates of the level that would be neutral for the economy" puts the level of the real neutral rate in the United States at about 0.5 percent (Powell 2018). In Japan, which has been faced with very low neutral rates for a long time, the central bank has engaged in aggressive monetary easing, including directly targeting long-term interest rates (Kuroda 2016). Similarly, European policymakers have highlighted the equilibrium rate of interest as the key policy variable (Constâncio 2016; Draghi 2016), while the recent paper from the European Central Bank (ECB 2018, 5) concluded that "most of our estimates of R^* for the euro area have been negative regardless of the type of model used."

Our main contribution in this paper is to recognize that the neutral real interest rate is not a deep structural feature of an economy but instead reflects both how it is embedded in the global economy and how fiscal policy is set. The neutral interest rate for an individual open economy will depend on its current account position, which in turn depends on its real exchange rate, which is itself a function of current and prospective real interest rates. It is therefore hard to interpret estimates of the neutral interest rate for a single open economy. We therefore estimate the neutral real rate for the industrial economies taken in the aggregate. We show that our aggregate can to a good approximation be thought of as a closed economy. Our estimates suggest that the advanced economies' neutral real rate has declined by over 300 basis points since 1980 and is now in the neighborhood of zero.

We emphasize that this significant decline would have been substantially greater except for the buildup of government deficits and debt over the last generation and the increasing generosity of social insurance programs, particularly increases in old-age pensions. Although the uncertainties inherent in any calculation are enormous, we estimate that, with constant fiscal and social insurance policies, neutral real interest rates would have declined by about 700 basis points and would now be very substantially negative. Equivalently, our estimate is that, with constant real interest rates, the gap between private saving and private investment rates in the industrialized world has widened by over 10 percent of GDP.

Our findings raise the possibility that the industrialized world as a whole will increasingly in the years ahead face the challenges that Japan has faced over the last decade. The emergence of a large gap between private saving and private investment at normal interest rates is the essence of secular stagnation. We believe that it has profound implications for macroeconomics, pointing to the need for a "new Keynesian economics" that bears a substantial resemblance to the old Keynesian economics of the 1950s and 1960s. It suggests the need for substantial revisions in the conventional wisdom regarding monetary policy based on inflation targeting, the role of fiscal policy in macroeconomic stabilization, and the appropriate level of budget deficits and possibly social insurance and regulatory policies.

We make two methodological choices in this paper. First, we show that the current account balance of the advanced economies, taken as a whole, has been small and stable over the past four decades; and, given this, we argue that it is preferable to view the advanced economies as a fully integrated bloc—a departure from the literature that tends to focus on individual countries when estimating neutral real rates. Second, we show that the dominant force driving the downward trend in real rates is common to a wide range of asset classes with differing characteristics. This explains the focus in the paper on forces driving the balance of desired saving and investment, as opposed to those that relate to the liquidity or safety attributes of any particular asset class.

To set the scene, we present the results from the econometric exercise estimating R^* for the industrialized world as a whole, which are that the advanced economies' neutral rate—which we call AE R^* for brevity—has declined by about 300 basis points over the past half century. This large decline in the relative price of consumption today versus consumption tomorrow has meant that the observed saving and investment ratios remained broadly stable. In other words, the large decline in R^* had been a symptom of the excess saving problem, and has masked the underlying shifts in desired saving and investment propensities. To illustrate the magnitude of this problem, we calculate a counterfactual gap between saving and investment propensities under the hypothetical scenario of a constant interest rate. Our calculations suggest that, if interest rates had not declined, the excess saving gap in the advanced economies taken together would be very large—likely, north of 10 percent of GDP. In the remainder of the paper, we study the various factors that underlie this phenomenon.

Our main contribution is the analysis of public policies and their impact on R^*. We show that all the available evidence points to a sizable positive influence: that the secular trends in public policies in the industrialized world have helped to reduce the excess saving problem.

Policies may affect the interest rate through a range of channels. We review these mechanisms, focusing on the role of government borrowing, which is the main focus of both theoretical and empirical literatures in macroeconomics. We then survey the existing empirical estimates of the impact of government debt on interest rates. Simple calculations using observed estimates of the impact of deficits on interest rates suggest that the increase from 18 percent to 68 percent in the public debt-to-GDP ratio of the advanced economies should, ceteris paribus, have raised real rates by between 1.5 and 2 percentage points over the last four decades. A similar calculation based on the existing empirical literature on the link between Social Security and private saving suggests that the increase of about 3 percentage points in Social Security spending to GDP may have increased interest rates by a further 50–100 basis points.[1] Increasing old-age health expenditures likely have had a further positive impact. This analysis leads to the conclusion that the fall in real long-term interest rate observed in the data masks an even more dramatic decline in the equilibrium "private sector" real rate.

To build further understanding of the mechanisms involved and to cross-check the magnitudes of these effects, we study these phenomena in a dynamic general equilibrium framework. We construct two tractable models, each one designed to capture different channels through which policies play out in equilibrium.

Building on the work of Mark Gertler (1999), the first model captures life-cycle behavior, with workers saving for retirement and retirees decumulating their wealth. Ricardian equivalence—the proposition that government borrowing decisions are neutral in equilibrium—does not hold in our model, making the effects of a range of government policies on real rates nontrivial. Specifically, after a change in government finances, there is some Ricardian offset, but unlike in the representative agent model, this offset is incomplete. We simulate the model with the profiles of government debt, government spending, Social Security, and old-age health care expenditures that match the experience of developed economies over the past 40 years. These simulations suggest that shifts in these policies have pushed equilibrium real rates up by about 3 percentage points between the early 1970s and today.[2]

Our second model focuses on individual risks and precautionary behavior, channels that are absent from the life-cycle model. When people cannot fully insure against future uncertainty surrounding their individual income, they value holding financial assets such as government debt for the purposes of self-insurance. This demand for assets is, in part, satisfied by governments issuing debt; and the more debt that is issued, the lower is its price and the higher is the interest rate. Our numerical explorations suggest that the increase in the supply of government bonds has pushed interest rates up by about 40–70 basis points through this channel.[3] Overall, then, we find that public policies may have pushed interest rates up by about 3.5–4.0 percentage points.[4]

1. For descriptions of the calculations that yield these numbers, see the end of section IV.
2. This number is derived from 3.6 – 0.4 = 3.2, rounded to 3, in the incomplete markets model shown in table 7 below.
3. This range reflects the second and third columns of table 7 below.
4. This refers to the italicized 3.6 in the middle column of table 7 below.

In the final section of the paper, we validate our models by using them to quantify the impact of some of the private sector forces that the existing literature suggested have been important in driving the decline in the neutral real interest rates. Specifically, we show that the estimates of the impact of the decline in expected future growth, the demographic shifts, and the rise in income inequality on neutral real rates are well within the ranges of estimates found by other researchers. This lends credibility to the core contribution of our paper, namely, the quantification of the boost that the public sector gave to neutral real rates in the advanced economies.

The remainder of this paper is structured as follows. Section I discusses two methodological issues underlying our analysis. Section II contains the results of the estimation of the long-term equilibrium real interest rate for the advanced economies. Section III starts with a discussion of the channels through which government policy influences the equilibrium rate; it then summarizes the results from the existing empirical literature that estimates the size of these effects; and finally, it uses these elasticities to calculate several back-of-the-envelope measures of how policies have affected AE R^*. In section IV, we set up the two general equilibrium models and use them to study the impact of government policies. Section V validates the models by using them to assess the impact of secular demographic changes, the slowdown in technology, and the rise in inequality. Section VI concludes.

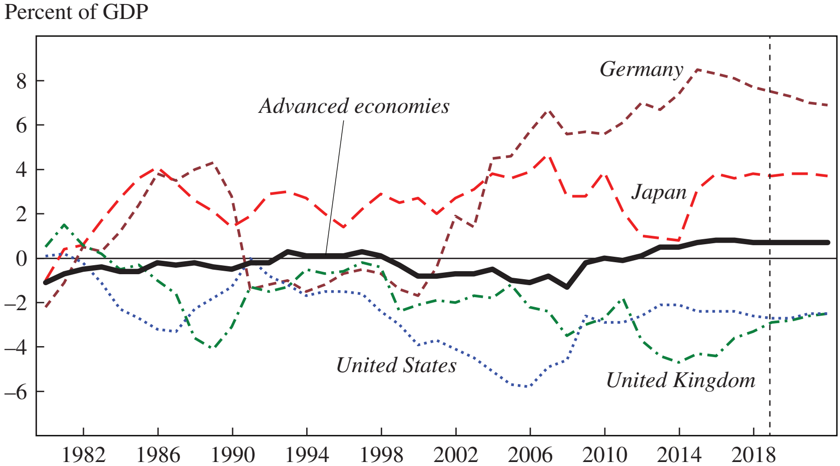
I. Understanding Neutral Real Interest Rates

We begin with a discussion of two methodological choices that permeate the analysis in this paper and where our approach differs from that of some other studies. The first is our treatment of the advanced economies as a bloc rather than focusing on individual countries. The second is our view that the decline in neutral real interest rates can be understood through the balance between desired saving and investment. This view leads us to focus on the macroeconomic forces affecting a broad range of returns, rather than on factors driving spreads or premia on particular financial instruments, and to deemphasize the importance of "safe asset shortage" theories for understanding the broad low frequency movements in interest rates.

I.A. The Advanced Economies as a Bloc

Our analysis assumes that the advanced economies bloc is fully integrated. In practice, we use aggregated data for all the developed countries (that is, the members of the Organization for Economic Cooperation and

Figure 1. The Aggregated Current Account of the Advanced Economies Bloc, and of Selected Individual Economies, 1980–2022[a]



Source: International Monetary Fund data.

a. The heavy solid line shows the current account for the advanced economies as defined by the International Monetary Fund.

Development, the OECD, whenever data are available), "as if" the bloc were a single economic entity. We treat this bloc as a large, closed economy with perfect internal capital mobility.

The perfect internal capital mobility assumption is justified by very large gross and substantial net capital flows between developed economies, strong commonality in trends in long-term real rates observed in the data, and high correlations in the performance of stock markets (Claessens and Kose 2017).

The assumption that the industrial economies as a whole can be treated as a closed economy is justified by the observation that their aggregate current account balance has fluctuated by less than 1.5 percent of GDP over the last 40 years (figure 1). Note also that the recent trend has been upward, suggesting that international capital flows have if anything operated to raise interest rates over time.

More importantly, our approach avoids the erroneous assumption implicit in much of the country-level analysis that the economies under consideration are closed. Current account balances for individual economies are large and variable; they are endogenous outcomes of the saving and investment propensities within each economy relative to the global average. For example, a country that runs a chronic trade surplus will be found to have a neutral real rate at a level where domestic demand is short of potential output, and the reverse will be true for a country running a chronic trade deficit. External balances should therefore be taken into account in such country-level analyses. We instead posit that the developed economies, taken together, experience structural excess saving, reflected in the trend decline in real interest rates without a discernible trend in their current account. At this level of aggregation, the country-level differences wash out, and econometric and theoretical analyses based on a closed-economy assumption are more credible.

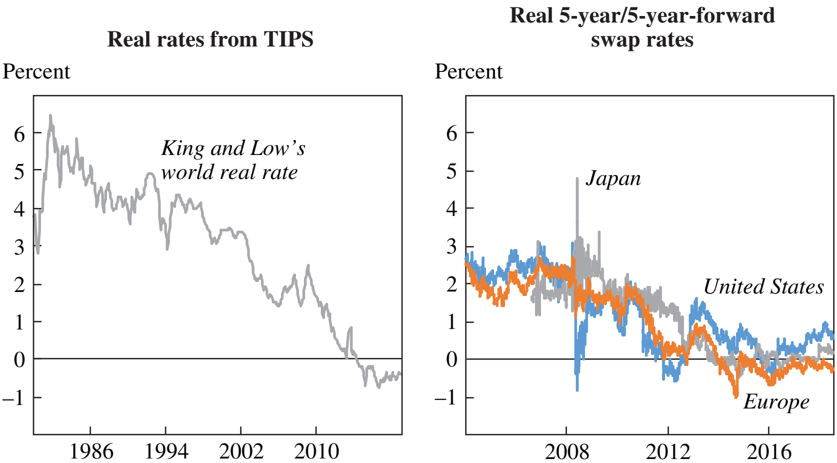
I.B. Excess of Desired Saving over Investment, and the Role of the Safety and Liquidity Premium

We carry out our analysis on the basis of the premise that, for analyzing long-term trend movements in neutral real rates, it is appropriate to focus on factors relating to saving and investment propensities rather than issues of liquidity or risk. Consequently, our analysis abstracts from aggregate uncertainty and differing levels of liquidity of various assets.

Several facts support our approach. First, the decline in rates on highly liquid securities tracks declines in yields on relatively illiquid government-indexed bonds and real swaps (figure 2), suggesting that the liquidity characteristics of government bonds play only a secondary role. Second, even in the United States, there has been little trend movement in spreads between Treasury securities and corporate securities in given rating classes, and though the pickup in equity risk premia has been somewhat more pronounced, it is nonetheless small relative to the decline in real interest rates over the decades (the left panel of figure 3). In any case, it is not clear whether one should interpret any changes in spreads as driven by changes in risk preferences or rather a result of changes in how risky the underlying assets are perceived to be. For instance, the recent global financial crisis has likely led to a reassessment of what it means that an asset is rated AAA, whereas the dot-com bubble appears to have had a lasting impact on the pricing of equities.

To get a sense of the relative importance of the trend decline in real returns versus changes in the dispersion between them, we summarize the patterns in the U.S. data using principal components analysis (PCA), which is a statistical procedure that summarizes the information in the correlated data series with a smaller set of mutually uncorrelated variables. The components are ordered in such a way that the first explains the highest share of variance in the data. The PCA thus offers a way of quantitatively distinguishing between the excess saving story, which drives the common

Figure 2. Real Interest Rates Estimated from Inflation-Linked Bonds and Forward Swaps, 1980–2022[a]



Sources: DataStream; Federal Reserve Bank of Saint Louis; Bloomberg; King and Low (2014).

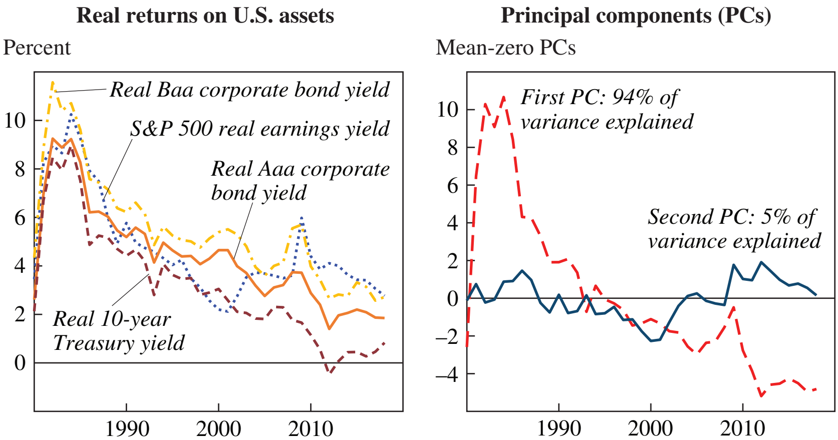
a. TIPS = Treasury Inflation-Protected Securities. The world real rate is calculated using the methodology of King and Low (2014): it is the average of interest rates on inflation-protected government debt securities across the Group of Seven, excluding Italy. Data are from DataStream and from an unbalanced panel. In particular, the figure relies on the U.K. inflation-indexed gilts in the early part of the sample. The U.S. TIPS yield is the yield on a constant maturity, 10-year Treasury Inflation-Indexed Security, retrieved from the FRED database of the Federal Reserve Bank of Saint Louis (code DFII10). Swaps data are from Bloomberg.

trend across all real rates, and the safety and liquidity story, which drives the dispersion between them.

When we perform this PCA on the set of U.S. real yields that span government debt, corporate bond and equity markets, we obtain results that are telling: the first principal component, which picks up the downward trend visible in all returns, explains 94 percent of the total variance in the underlying series (the right panel of figure 3). The second principal component, which appears to be related to the increase in the "convenience yield," explains only 5 percent. The very large share of total variance in the data accounted for by the common downward trend supports our focus in this paper.

This focus is also consistent with the finance literature that investigates the decline in neutral real interest rates in the presence of term and liquidity premia (Christensen and Rudebusch 2019; D'Amico, Kim, and Wei 2018). The "safe asset" literature finds a somewhat larger role for the convenience yield; but even there, the magnitudes are generally rather small relative

Figure 3. Real Returns in the United States: Data and Principal Components, 1980–2022[a]



Sources: FRED database of the Federal Reserve Bank of Saint Louis; Robert Shiller; authors' calculations.

a. The 10-year Treasury real yield post-2004 is the yield on 10-Year Treasury, Inflation-Indexed Security (code: DFII10). Before 2004, it is the nominal yield (GS10) minus inflation expectations measure from the Michigan survey (code: MICH). The real corporate yields are nominal yields (codes: Aaa, Baa), minus inflation expectations from Treasury Inflation-Protected Securities (post-2004) or from the Michigan survey (pre-2004). The real earnings yield is the inverse of the cyclically adjusted, total return price real earnings ratio from Robert Shiller (http://www.econ.yale.edu/~shiller/data.htm).

to the large trend decline in real rates. For example, using very different approaches, Marco Del Negro and others (2018) and Rachel and Thomas Smith (2017) concluded that the rise in the spread between risky and risk-free rates accounted for about 70 basis points of the decline in risk-free rates. This is less than a quarter of the overall decline in real neutral rates since 1980.

In summary, much of the available evidence points to a common underlying decline in real interest rates across different financial assets. This suggests that saving and investment propensities and how they have changed over time are the dominant underlying drivers of such a trend.

II. Estimating the Advanced Economies' Equilibrium Real Interest Rate

In this section, we estimate the natural rate of interest for the advanced economies, adopting what is perhaps the most celebrated applied empirical model designed for this purpose, which was originally developed by

Thomas Laubach and John Williams (2003) (hereafter, the Laubach-Williams model) and has recently been reapplied internationally by Kathryn Holston, Laubach, and Williams (2017b). Conceptually, this approach draws on two strands of the literature. By following Wicksell's (1898) definition of the natural rate as the rate consistent with stable inflation and output remaining at an equilibrium ("potential") level, it is well aligned with modern monetary theory, as presented by Carl Walsh (1998), Michael Woodford (2003), and Galí (2008). This literature is primarily concerned with fluctuations at the business-cycle frequency, where shocks move the economy around a stable steady state. In addition to these business-cycle shocks, the framework employed here is flexible enough to capture secular forces that affect the steady state.

II.A. A Sketch of the Model and the Estimation Procedure

Our approach to estimating the Laubach-Williams model is deliberately off the shelf: we use exactly the same procedures as the recent papers in that literature. Out contribution is solely to perform this exercise on the bloc of advanced economies as a whole. As such, we do not take a stance on the performance of the model, although we discuss some of the issues below.

The philosophy of the Laubach-Williams method is that the natural rate of interest is an endogenous object determined in general equilibrium, and as such it will depend on a host of socioeconomic forces, such as trends in preferences, technology, demography, and policies and policy frameworks. It is impossible to know and measure all the relevant factors. At the same time, a robust prediction of most workhorse macroeconomic models is that the natural rate should vary together with the economy's expected future trend growth rate.[5] To reflect the dependence on growth and on a range of possibly unknown other factors, the Laubach-Williams model assumes that the natural rate, denoted r_t^*, depends on the estimated trend growth rate of potential output g_t and a time-varying unobserved component z_t that captures the effects of other unspecified influences:

$$r_t^* = g_t + z_t. \tag{1}$$

The model further assumes that both the growth rate g_t and the unobserved component z_t are random walk processes:

$$g_t = g_{t-1} + \epsilon_{g,t} \qquad \epsilon_g \sim N\left(0, \sigma_g^2\right) \text{ and} \tag{2}$$

5. We discuss the rationale for this link in some detail in section IV.

$$z_t = z_{t-1} + \epsilon_{z,t} \qquad \epsilon_z \sim N(0, \sigma_z^2). \tag{3}$$

The model specification also admits shocks to the level of potential output. Denoting by y_t^* the natural logarithm of potential output at time t,

$$y_t^* = y_{t-1}^* + g_{t-1} + \epsilon_{y^*,t} \qquad \epsilon_{y^*} \sim N(0, \sigma_{y^*}^2). \tag{4}$$

In short, the Laubach-Williams model views the natural rate as the sum of two independent random walks. To achieve identification, Laubach and Williams add two further equations to the model. First, they specify a simple reduced-form equation relating output gap to its own lags, a moving average of the lagged real funds rate gap, and a serially uncorrelated error:

$$y_t = y_t^* + a_1(y_{t-1} - y_{t-1}^*) + a_2(y_{t-2} - y_{t-2}^*) + \frac{a_r}{2}\sum_{j=1}^{2}(r_{t-j} - r_{t-j}^*) + \epsilon_{y,t} \tag{5}$$

$$\epsilon_y \sim N(0, \sigma_y^2).$$

The key in this estimated investment-saving relation is the a_r coefficient, which we expect to be negative. Second, Laubach and Williams add the reduced-form Phillips curve to the model, linking current inflation, π_t, to lagged inflation and the output gap:

$$\pi_t = b_\pi \pi_{t-1} + (1 - b_\pi)\pi_{t+2,4} + b_y(y_{t-1} - y_{t-1}^*) + \epsilon_{\pi,t} \tag{6}$$

$$\epsilon_\pi \sim N(0, \sigma_\pi^2),$$

where the standard theory would suggest that coefficient b_y is positive.

The system presented above can be written in a state-space form, and the Kalman Filter can be used to estimate the unobservable states. To estimate the model, we use data for the advanced economies as a bloc. The data comprise the log of quarterly real GDP, core inflation, and long-term interest rates over the period 1971:Q1–2017:Q4 for the aggregated sample of OECD countries. The interest rate series is the average of long-term nominal interest rates across an unbalanced panel of 36 OECD economies.[6] To calculate real rates, we subtract from nominal rates a simple measure of expected inflation, constructed as the moving average of past core inflation

6. The results are robust to using weighted average or median of the interest rates across countries. Given the strong co-movement, these interest rate series are close to each other.

Table 1. State-Space Model Parameter Estimates

Parameter point estimates (t *statistics in parentheses)*							
a_1	a_2	a_r	b_π	b_y	σ_y	σ_g	σ_z
1.71	–0.79	–0.04	0.90	0.09	0.25	1.03	0.31
(21.65)	(10.28)	(2.3)	(17.78)	(2.06)	(5.30)	(29.63)	(9.38)

Average standard errors around the estimates		
y^*	r^*	g
1.19	3.12	0.16

Source: Authors' calculations.

rates, in line with the work of Holston, Laubach, and Williams (2017b). See online appendix A for further details on the data and the estimation procedure.[7]

II.B. Results

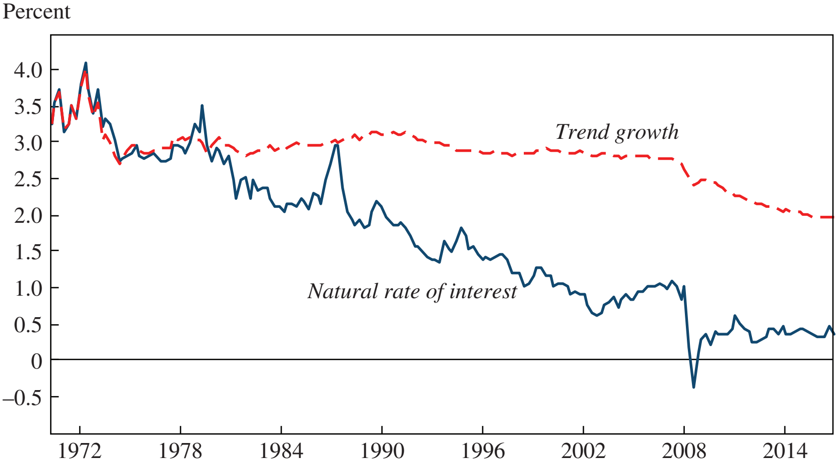
Table 1 shows the coefficients of the estimated model. The point estimates are all significantly different from zero and have expected signs. In particular, a positive interest rate gap reduces the output gap, while a positive output gap raises inflation. Table 1 also shows the standard errors around the estimated trends, which are large, especially those around the estimates of the equilibrium real rate. These wide standard error bands are not specific to our results—indeed, they are a norm in the literature. For instance, Holston, Laubach, and Williams (2017b) report similarly large errors for individual economies. These errors are, to an extent, an artifact of the long sample, as they reflect the cumulative uncertainty of the underlying drivers of equilibrium rates. Nonetheless, these large error bands should act as a reminder of the high uncertainty surrounding the econometric estimates of equilibrium interest rates.

Figure 4 contains the key results. According to our estimates, AE *R** declined steadily from the 1980s onward, and fell sharply during the 2008 global financial crisis.[8] It then stabilized at low levels (0.5 percent). The estimated growth rate of potential output was broadly stable up until the crisis, and declined during the crisis by about 1 percentage point. Thus,

7. The online appendixes for this and all other papers in this volume may be found at the *Brookings Papers* web page, www.brookings.edu/bpea, under "Past BPEA Editions."

8. Estimates for the first decade should be taken with a grain of salt, because the model is less accurate during the first few years of the sample when the initial conditions play a larger role.

Figure 4. Changes in AE R^* and Trend Growth, 1971–2016



Source: Authors' calculations.

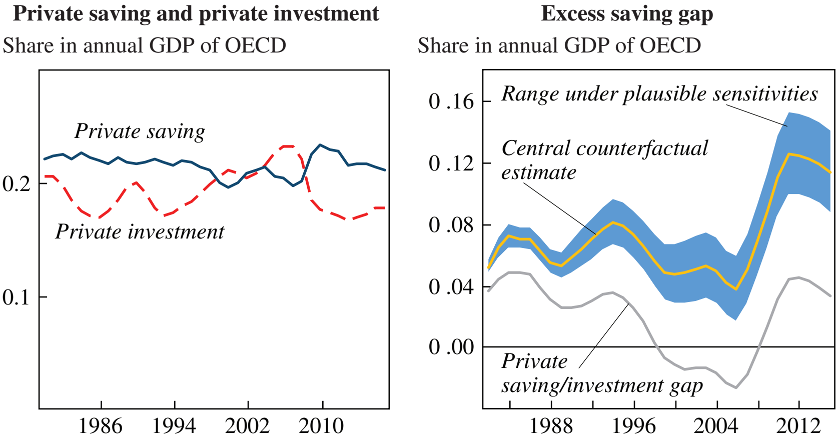
the model suggests that the bulk of the decline in real interest rates is due to factors other than trend GDP growth. This is consistent with the literature that finds only a loose connection between actual GDP growth and interest rates in the historical data (Hamilton and others 2016).

These results corroborate other existing findings in the literature. In particular, Holston, Laubach, and Williams (2017b) estimated declines in real rates for the United States, Canada, the euro area, and the United Kingdom of about 2.3 percentage points between 1990 and 2017; for comparison, the decline over this period for the advanced economies as a whole that we estimate here is about 2 percentage points.

Overall, despite high uncertainty surrounding the point estimates of these trends, we interpret the results of this exercise as broadly in line with the country-level findings in the literature. Indeed, given the high level of aggregation, we find it encouraging that the estimated unobservables do well at picking up the main events, such as the global financial crisis, during which our estimate of AE R^* declines very sharply.

Perhaps more significantly, our estimates of the decline in the neutral real rate track the evolution of 10-year real yields depicted in figure 2. This both provides further corroboration of our estimates and suggests a market judgment that real rates are likely to remain low for the foreseeable future.

Figure 5. Private Saving and Private Investment in Advanced Economies: The Level and Gap, 1980–2014/16[a]



Sources: International Monetary Fund; Organization for Economic Cooperation and Development; authors' calculations.

a. The left panel of this figure shows purchasing power parity–weighted gross private saving and gross private fixed-capital formation across the following countries: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Japan, South Korea, Latvia, Lithuania, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, the United Kingdom, and the United States. The figure on the right shows the result of a simple counterfactual exercise where we calculate the private saving/investment gap under a scenario of no decline in the long-term interest rate since the 1980s. The swath contains the counterfactual for values of responsiveness of saving/investment ratios to interest rates between 2 and 4, and the central counterfactual estimate assumes the sensitivity of 3.

II.C. The Fall in AE R* *and the Excess Saving Problem*

The decline in the neutral real rate of this magnitude is a symptom of deep, fundamental changes that have taken place in the developed economies over the last half century. A useful way to think about these trends is through the lens of desired saving and investment, with the desire to save running ahead of the desire to invest. However, illustrating the fundamental change in this space is not straightforward, because the large fall in the intertemporal price—the interest rate—meant that the *observed* saving and investment ratios remained broadly stable throughout this period. The left panel of figure 5 shows the realized purchasing power parity–weighted private saving and investment ratios in the OECD, in proportion to the aggregate GDP of the OECD. The saving ratio is almost completely stable, and though there is some movement in the ratio of private investment to GDP, there certainly is no strong trend.

To assess the magnitude of the forces that operated under the surface in terms of excess saving over investment, one needs to perform a counterfactual analysis. Here we present a simple but telling attempt. Specifically, we calculate the counterfactual difference between private sector saving and investment—the counterfactual private sector saving/investment gap—under an assumption of no decrease in the interest rate since the 1980s. To construct such a counterfactual, we need an estimate of the strength of the link between desired saving and investment and the interest rate. We rely on the estimates reported in empirical literature,[9] which suggests that the elasticity of desired saving is in the region of 0.3 to 0.7 and the elasticity of desired investment is about –0.5 to –0.7. With average saving- and investment-to-GDP ratios at about 20 percent, elasticities of this magnitude suggest that a decline of 1 percentage point in the real interest rate is associated with a widening of the saving/investment gap of between 2 and 4 percentage points, with the central view of the sensitivity of about 3.[10] Given the uncertainties, we report the counterfactual gap under this range of sensitivities.

The main message from these simple calculations is striking: absent the cushioning decline of the interest rate, the excess saving gap would have been very large: the right panel of figure 5 indicates that it would have been between 9 and 14 percentage points.

Motivated by the size of these movements, we now turn to the discussion and analysis of the forces behind them. Our contribution is the focus on the role that public policies have played over this period.

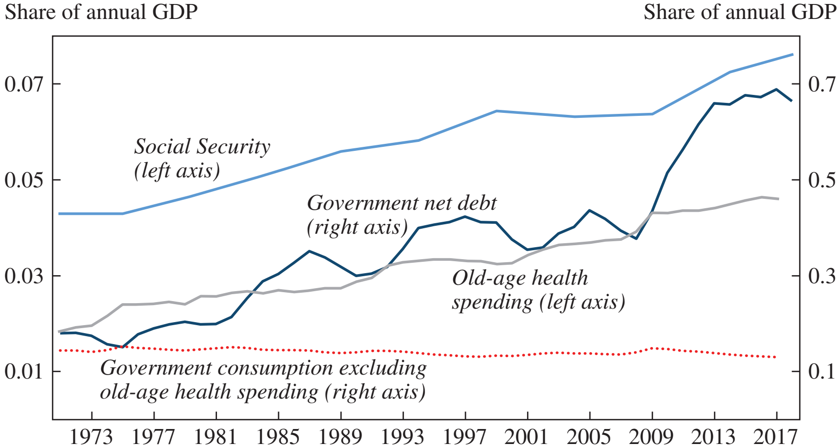
III. Government Policy and the Equilibrium Interest Rate

Over the past several decades, government policy in the developed world has shifted significantly in at least four respects (figure 6). First, government debt has risen, from about 20 percent of GDP to about 70 percent (government consumption—excluding health care—remained relatively stable). Second, old-age payments administered through the Social Security and health care systems have gone up, respectively, from about 4 percent

9. See Rachel and Smith (2015) for a review.

10. The central estimate given in the right panel of figure 5 errs on the side of caution, assuming the sensitivities at the lower end of these ranges. When interest rates are 5 percent, a decline of 1 percentage point in the interest rate constitutes a 20 percent decline. Given an elasticity of 0.3 of saving and –0.5 of investment, this is associated with a change in desired saving of 6 percent and investment of 10 percent, which add up to about 3 percent of GDP when these ratios are about 20 percent of GDP.

Figure 6. The Advanced Economies' Government Policy Ratios (in Proportion to GDP), 1971–2017[a]



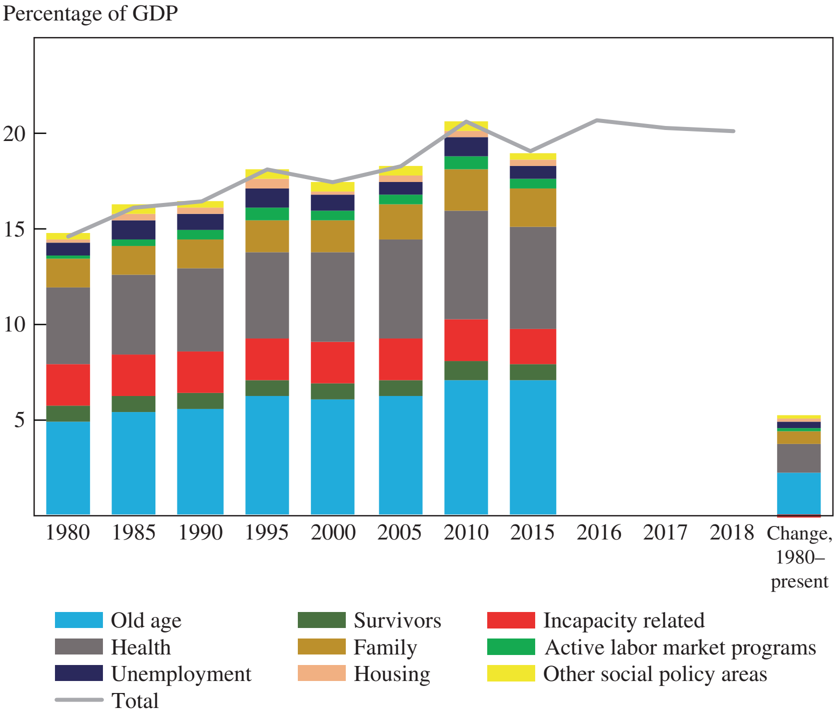
Source: OECD data.

a. This figure shows OECD aggregates in proportion to total GDP. The government net debt line measures general government net financial liabilities, from the *OECD Economic Outlook Database*. It includes the net government debt held by the public, and also other net liabilities of the government. For example, in the United States in 2017, net financial liabilities as reported by the OECD were 80 percent of GDP, while net debt held by the public was 75 percent of GDP. Government consumption figures represent general government final consumption expenditures, adjusted by subtracting old-age health spending (note that this series excludes the Social Security transfers by default). Old-age health spending is calculated as the aggregate health spending on those age 65 and above. The overall health spending figures are from OECD / World Health Organization statistics on sources of funding for health care. They include health care financed directly by the government and from compulsory schemes. The old-age share is then calculated under the assumption that 60 percent of total health spending is directed at the older demographic groups, which is consistent with the evidence available for a handful of OECD countries.

to about 7 percent, and from about 2 percent to about 5 percent of GDP, accounting for the lion's share of the increase in total social spending (figure 7). Third, significant changes have taken place in tax policies. The effective corporate tax rates in the high-income economies have fallen, from about 32 percent at the turn of the century to 24 percent more recently.[11] Wealth taxes, which were operational in 12 OECD countries in 1990, remain in place only in 4 countries today (OECD 2018). And, as documented by Thomas Piketty and Emmanuel Saez (2007), the overall progressivity of the tax system has decreased in some jurisdictions—notably, in the United States and the United Kingdom.

11. For the details, see OECD (2019).

Figure 7. Public Social Spending in the OECD Countries, 1980–2018[a]



Source: OECD Social Expenditure Database (SOCX).
a. The data for 2016–18 are preliminary and the breakdown by function is not yet available.

These shifts are likely to have had a profound impact on the economy in general, and on the equilibrium rate of interest in particular. Specifically, all these shifts—perhaps with the exception of tax changes—are likely to have pushed interest rates higher over the past 30 years. In this and the next section, we turn to the analysis of the impact of these policy shifts on the natural rate, with the ultimate goal to inform the counterfactual "pure" R^* that would prevail without government intervention.

We focus on government debt, Social Security, and health care spending, leaving the formal analysis of the impact of tax changes for future research. We find that shifts in government policy have likely pushed equilibrium rates of interest up by a significant amount over the period in question. As a rough rule of thumb for the magnitudes involved, our analysis suggests that the tripling of the government debt over the past half century has raised rates by 1.5 percentage points, while the expansion of social spending of about 5 percent of GDP has added a further 2.5 percentage points.[12] Although the precise magnitudes of these multipliers are subject to substantial model and statistical uncertainty, the qualitative conclusion is clear: If public policy had not responded, the advanced world's equilibrium rate would likely be deeply negative.[13]

III.A. A Brief Review of the Theoretical Arguments

We begin by reviewing the effects of government policy on the equilibrium interest rate, focusing on government borrowing, given that this has been the main subject of the large body of literature in macroeconomics on which we can draw.[14]

In the canonical neoclassical model with complete markets and infinitely lived agents, Ricardian Equivalence holds, and neither the deficit nor the debt is relevant, because the representative household can fully offset the changes to the government's borrowing policy through its saving decisions. Thus, independent shocks to government borrowing alone have no effect on the equilibrium interest rate. The neoclassical model instead emphasizes the link between the stock of capital and the interest rate: in equilibrium $r = f'(k) - \delta$.[15] Thus, government policies affect the interest rate only to the extent that they have an impact on the stock of private capital.

However, in the micro-founded modern macroeconomic models that depart from the representative agent and complete markets assumptions, Ricardian Equivalence does not hold, and government transfer policies affect the equilibrium allocations through several distinct channels.

12. This rounds the numbers in table 7 below. The 1.5 refers to the sum of the second column, second and third rows (1.2, but rounded to 1.5 here). The 2.5 rounds the sum of the fifth and sixth rows of the second column (1.2 + 1.1).

13. A corollary of this link between government debt and interest rates is that a higher value of public debt, compared with market expectations, is likely to raise the natural interest rate. For an analysis of this argument, see Kocherlakota (2015).

14. In our work, we do not explicitly model the impact of quantitative easing (QE) policies. One kind of QE encompasses policies that swap risky assets for safe assets and includes the Federal Reserve's initial round of QE (QE1) in the United States or the Long-Term Refinancing Operation (LTRO) in the euro zone. Such a policy may raise the short-run rate (Caballero and Farhi 2018), whereas we focus on the long-run rate. Another kind is a policy whereby the central bank issues reserves to buy risk-free debt. Such a policy is primarily a maturity transformation of government debt, rather than a change in the total availability of investable assets.

15. On the balanced growth path, the level of effective capital stock adjusts such that the interest rate simultaneously satisfies the balanced growth version of the representative household's Euler Equation: $r = \frac{1}{IES} \cdot g + \theta$, where *IES* is the intertemporal elasticity of substitution and θ is the rate of time preference.

First, the intertemporal transfers—that is, *redistribution across time*—matters if peoples' planning horizons are finite. This could be because of finite lives coupled with a less-than-perfect bequest motive, as in the seminal models of Peter Diamond (1965) and Olivier Blanchard (1985), or perhaps due to the time-dependent preferences and myopic behavior pioneered by David Laibson (1997). The reason is intuitive: with finite planning horizon, agents currently alive expect to shoulder only a part of the financing burden that comes with today's transfer; the rest is to be serviced by future generations. Such transfers thus affect agents' wealth and their consumption and saving plans.

Second, *transfers across agents* can affect aggregate consumption and saving (and hence the interest rate) if agents have different marginal propensities to consume (MPCs). Differences in MPCs could arise because of several distinct features of the economic environment. They could be a result of uninsurable risks and binding borrowing constraints, as in the works of Rao Aiyagari (1993), Aiyagari and Ellen McGrattan (1998), and the model of Hyunseung Oh and Ricardo Reis (2012). They could emerge because some agents have little to no liquid wealth, preventing them from adjusting their consumption, as in the paper by Greg Kaplan, Giovanni Violante, and Justin Weidner (2014). Another reason may be the life cycle: the propensity to consume may differ between workers and retirees, as shown by Gertler (1999); or it may vary with age, as shown by Etienne Gagnon, Benjamin Johannsen, and David Lopez-Salido (2016) and by Gauti Eggertsson, Neil Mehrotra, and Jacob Robbins (2019). Heterogeneous MPCs and distortionary taxes deliver this result in the savers-spenders model of N. Gregory Mankiw (2000).[16] In all these models, government transfers from a low-MPC agent to a high-MPC agent will boost the aggregate desire to consume and lower desired saving, thereby raising the interest rate.

The third way in which government policy affects interest rates is what may be called a *precautionary saving channel*. One facet of this channel is that government policies can directly reduce the risks faced by the agents. The mechanism is close to the one analyzed by Eric Engen and Jonathan Gruber (2001). Under imperfect insurance, agents who face idiosyncratic risks—for example, those related to health or unemployment—attempt to

16. In the savers-spenders model of Mankiw (2000), if taxes are levied lump-sum, a deficit-financed transfer that permanently increases the level of debt does not affect the stock of capital or the interest rate in the long run. The reason is that the interest rate is pinned down by the savers, who are infinitely lived and Ricardian.

self-insure through saving. This precautionary saving motive acts to push the interest rate below the rate that would prevail in a complete market economy (where all risks are insurable and thus do not affect the agents' behavior). Government policies such as social insurance will affect the importance of precautionary saving: a stronger social safety net or higher unemployment and disability benefits curtail the associated risks, curbing the desire to save. Conversely, the lack of social insurance means that agents need to rely on their own resources when experiencing hardship, making personal saving a priority. However, as illustrated in figure 7, the overall size of the social safety net across the OECD has changed little over the period in question. We do not attempt to model it here, but leave it as an important direction to be explored in future research.

The other facet of the precautionary saving channel—and the one we focus on in this paper—works through the provision of assets that agents use to insure themselves against shocks. This mechanism is at the heart of the work of Aiyagari and McGrattan (1998), and it has recently been discussed in the context of secular stagnation by Ricardo Caballero, Emmanuel Farhi, and Pierre-Olivier Gourinchas (2016) and by Caballero and Farhi (2018). The intuition we have in mind is simple: a rise in government debt raises the overall supply of assets in the economy, which, all else being equal, pushes interest rates up. Indeed, there is evidence in the data that government debt constitutes a nontrivial proportion of the total investable financial assets in the developed world, so that this channel can have a quantitative bite. The estimates of the share of government bonds in total financial assets range from one-third in the United States to two-thirds in Japan (Kay 2015).

In summary, macroeconomic theory developed over the past couple of decades has enriched the basic model of Frank Ramsey and Robert Barro (see Barro 1974), with several channels that make government policy a relevant determinant of the long-term interest rate. We now turn to the empirical evidence that has been accumulated in parallel to these theoretical advances.

III.B. Empirical Evidence on the Link between Government Debt and Long-Term Interest Rates

The main challenge when estimating the effect of government borrowing on interest rates is the large number of potentially confounding factors, which may make simple regressions of interest rates on debt spurious and uninformative. For example, deficits will tend to expand when the economy weakens, which is also the time when interest rates tend to fall.

This means that the simple regression coefficients are likely to be biased downward.

We shall not attempt a full-blown empirical assessment in this paper, and instead present a summary of the empirical estimates in the literature. For an interested reader, online appendix B illustrates several challenges in estimating the causal relationship between equilibrium interest rates and government debt through a simple empirical exercise for the United States, Canada, the euro area, and the United Kingdom. These challenges include the presence of international capital flows and of endogenous responsiveness of policy to an excess of private saving over private investment, both of which are likely to attenuate the individual-country estimates of the impact of deficits on interest rates. Instead, we present the estimates from a broad body of literature that attempted to deal with these and other confounding factors in finding the link between government finances and real rates.

Several key studies in the empirical literature have focused on the United States. In a chapter in the *Handbook of Macroeconomics* at the turn of the century, Douglas Elmendorf and Mankiw (1999) reviewed the theoretical and empirical literature on the Ricardian Equivalence proposition, and they concluded that, though the studies that attempted to estimate the impact of government finances on interest rates cannot reject the null hypothesis of zero impact, they suffer from lack of statistical power.[17] More recent work appears to be more conclusive. In their literature review, William Gale and Peter Orszag (2003) conclude that the effect of government deficit on real rates is positive and economically significant: an increase of 1 percentage point in the deficit-to-GDP ratio tends to raise interest rates by about 50–100 basis points. And the two most authoritative contributions on the topic suggest estimates that are significant, albeit somewhat smaller. Laubach (2009) studies how forward rates on government securities react to news in the fiscal forecasts of the Congressional Budget Office. The identifying assumption in his work is that long-term rates and forecasts are not contaminated by current events and shocks at the business cycle frequency. According to his estimates, a rise in the

17. They write of the literature that tends to find close to zero effect of government deficit on rates: “Our view is that this literature . . . is ultimately not very informative. . . . Plosser (1987) and Evans (1987) generally cannot reject the hypothesis that government spending, budget deficits, and monetary policy each have no effect on interest rates. Plosser (1987) also reports that expected inflation has no significant effect on nominal interest rates. These findings suggest that this framework has little power to measure the true effects of policy” (Elmendorf and Mankiw 1999, 1658).

government deficit of 1 percentage point of GDP raises interest rates by about 20–30 basis points; an equal increase in the debt-GDP ratio results in a rise of about 3–4 basis points. He asserts that these flow- and stock-multipliers are broadly consistent, because of the autocorrelation of the deficits observed in the data.[18] Another important contribution to this literature is that by Engen and Glenn Hubbard (2004), who consider a host of specifications linking interest rates or changes in interest rates to government debt or to the deficit, both contemporaneously and in a forward-looking setting. Their results suggest that a rise of 1 percentage point in government debt to GDP pushes interest rates up by about 3 basis points, broadly in line with Laubach's findings.[19]

Further evidence is available for the advanced economies beyond the United States. In an international setting, Anne-Marie Brook (2003) documents that the range of estimates of the effect of an increase of 1 percentage point increase in the government debt-to-GDP ratio on interest rates is 1–6 basis points, with the corresponding range for an increase of 1 percentage point in deficits in the region of 20–40 basis points. In an important study of the euro area, Riccardo Faini (2006) finds that a rise of 1 percentage point in deficits at the euro area level raises long-term rates by about 40 basis points, close to—and if anything, higher than—the U.S. multipliers. Considering an even wider panel of 19 OECD economies spanning the years 1971–2004, Noriaki Kinoshita (2006) finds that the effect of a rise of 1 percentage point in the government debt-to-GDP ratio is to raise interest rates by 4–5 basis points.

A complementary way to assess the size of these effects is to consider simulations from large-scale models used for quantitative analyses in policy institutions. Because these models are carefully estimated using real-world data, they should be able to provide a steer as to the size of the effects. A well-known example is the FRB/US model, which is used and maintained by researchers at the Federal Reserve Board (Laforte and Roberts 2014). In a recent speech, Stanley Fischer (2016) uses this model to estimate the impact of a persistent increase in the deficit on real rates, and finds that an increase of 1 percentage point in the deficit raises the equilibrium rate by between 40 and 50 basis points, depending on whether

18. Specifically, he estimates the autocorrelation of 0.83, implying that the 1 percentage point rise in the deficit should have 1/1 – 0.83 = 6 times the effect of a 1 percentage point rise in debt—broadly in line with what he finds.

19. The results vary across different specifications, highlighting that the precise econometric details matter for the conclusions of this line of empirical research.

Table 2. The Impact of Government Borrowing on the Interest Rate: A Summary of the Literature

Study	*Country or region*	*Impact of 1 percentage point increase in deficit-GDP ratio (basis points)*	*Impact of 1 percentage point increase in debt-GDP ratio (basis points)*
Gale and Orszag (2002)	United States	50–100	—
Laubach (2009)	United States	20–30	3–4
Engen and Hubbard (2004)	United States	18	3
FRB/US model	United States	40–50	—
Faini (2006)	Euro area	40	—
Brook (2003)	Advanced economies	20–40	1–6
Kinoshita (2006)	19 OECD economies	—	4–5
Average		*38*	*3.5*

Sources: The studies listed in the first column; information on the FRB/US model can be found at https://www.federalreserve.gov/econres/us-models-about.htm.

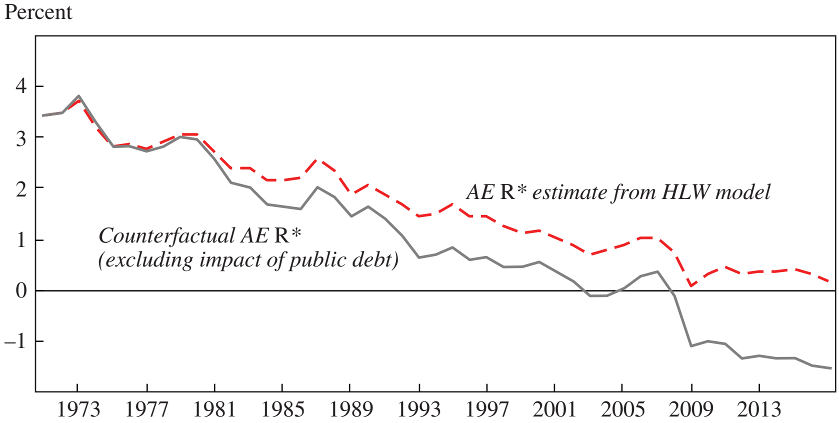
the deficit increased because of a tax cut (smaller effect) or a rise in government spending (larger effect). These figures are thus slightly larger than the empirical estimates cited above.

In summary, the estimates in the literature paint a fairly consistent picture: a rise of 1 percentage point in the deficit tends to raise interest rates by about 40 basis points; while a rise of 1 percentage point in the debt-GDP ratio results in an increase of about 3.5 basis points (table 2). We suspect this figure is an underestimate of the impact of an exogenous increase in budget deficits on real rates because fiscal expectations are measured with error, because any one country can import capital and so attenuate rate increases when budget deficits increase, and because there will be a tendency—as fiscal policy is used to stabilize the economy—for periods of low neutral real rates to coincide with periods of expansionary fiscal policy.

III.C. The Historical Impact of Government Borrowing on R^*

The elasticities identified in the empirical research combined with the historical path of government borrowing give simple back-of-the-envelope estimates of the historical influence of fiscal policy on real interest rates. Over the past 40 years, the increase in government debt in the OECD has likely pushed interest rates higher, perhaps by as much as 2 percentage points. The measure of R^* that excludes the impact of public debt has hovered around zero since the early 2000s, and remains negative at the moment (figure 8).

Figure 8. The Advanced Economies' R^*, Adjusted for the Impact of Government Debt, 1971–2017[a]



Source: Authors' calculations.

a. HLW = the model of Holston, Laubach, and Williams (2017a). This figure shows the estimated equilibrium real interest rate in the advanced economies, and an adjusted measure that subtracts the impact of government borrowing using the average elasticities reported in table 2.

III.D. The Link between Social Security and R^*

Social Security constitutes both an intertemporal and a between-group transfer. To be able to calculate by how much changes in Social Security have had an impact on the neutral real rate, we need the estimates of the impact of Social Security on individual saving, and also the differences in the MPCs of the groups funding and receiving the Social Security transfer.

A large body of literature has analyzed the first of these.[20] Several researchers relied on aggregate time series within a country. An example is the study by Martin Feldstein (1974), who finds a significant offset in the region of 30–50 percent of private saving to Social Security changes in the United States. However, other studies in this literature argued that the effects may be smaller and are highly uncertain. The second approach, based on micro data in the cross section of individuals, estimates a private sector offset of between 0 and 50 percent (Feldstein and Pellechio 1979). The cross-country studies find little to no effect (Barro and MacDonald 1979). More recent papers focus on pension system reforms to sharpen

20. See CBO (1998) for a review.

identification, and find significant responses to private saving (Attanasio and Brugiavini 2003; Attanasio and Rohwedder 2003). Overall, the literature is consistent with private saving reacting to the changes in Social Security, with elasticities between –0.3 and –0.4 representing the central tendency among a wide range of available estimates.

The impact of the between-group transfer depends on the differences in MPCs across the two groups: taxpayers and retirees. The traditional life-cycle logic suggests that retirees have a higher MPC relative to working-age individuals, although the evidence on the quantification of these differences is scarce. Christopher Carroll and others (2017) suggest that the difference could be in the region of 0.3.

Under these assumptions, the increase in Social Security of about 3 percentage points of GDP that we observed (figure 6) would have led to a decrease in desired saving of about 1 percentage point through the intertemporal channel and another 1 percentage point through the across-groups redistribution.[21] Based on the multipliers used in the calculations underlying figure 5, the overall decrease of 2 percentage points in desired private saving may have led to a rise in R^* of between 50 and 100 basis points. And the rise in old-age health care spending would have added further upward pressure on real rates.

To sum up, simple calculations suggest a very substantial upward impact of public policies on R^* over the past half century. To develop further intuition and to consider other mechanisms through which public policy may have affected the interest rate, we now turn to a complementary approach: a general equilibrium modeling framework.

IV. Government Policy and R^*: A Model-Based Approach

In subsection III.A, we outlined various channels through which government debt may affect the equilibrium real interest rate; our goal in this section is to illustrate their quantitative importance within a general equilibrium framework. We want our approach to be simple and transparent, providing a credible complement and a cross-check to the empirical analysis given above. To achieve these goals, we build two general equilibrium models: one capturing the finiteness of life and life-cycle heterogeneity, and another that focuses on precautionary behavior.

21. Both figures are obtained by multiplying the change in Social Security of 3 percentage points by 0.3.

IV.A. Two General Equilibrium Models

The first model, which builds closely on Gertler (1999), highlights life-cycle heterogeneity. In this economy, ex-ante identical individuals are at different points in their lives; some are working, and some have already retired. This drives the differences in their consumption and saving behavior. The framework is similar to that of Blanchard (1985) and Menahem Yaari (1965)—individuals face thr constant probability of death, and so their horizons are finite—but, in addition to their model, workers retire and finance consumption with saving until death.

The second model is a Bewley–Huggett–Aiyagari economy with incomplete markets and uninsurable income risk at the level of an individual household. A similar model was considered by Aiyagari and McGrattan (1998), who also studied the role of government debt in equilibrium allocation in the presence of idiosyncratic risk. The main differences between ours and their approach are that (1) we calibrate the risk component of the income process to deliver a realistic dose of uncertainty, which implies that distributions of income and assets in the model broadly match distributions observed in developed economies such as the United States;[22] and (2) we cast the model in continuous time, taking full advantage of the recent analytical and computational discoveries in macroeconomics.

Here, we sketch the main workings of the two models and develop the intuition; a more detailed description of the models is available in online appendix C for the life-cycle model and online appendix D for the incomplete market model.

IV.B. A Model of Finite Lives and Life-Cycle Heterogeneity

With respect to demographics and preferences, there are two stages of life: work and retirement, with exogenous transition probabilities. That is, each worker faces a given probability of retirement $1 - \omega$, and, once a retiree, a given probability of death $1 - \gamma$. Population grows at a gross rate $1 + n$.

There is no aggregate risk; the only sources of uncertainty facing an individual are the risk of retirement while a worker (associated with a loss of labor income) and the risk of death while a retiree. Left unchecked, these sources of risk would affect agents' behavior. This would make aggregation problematic, and, more important, it would be unrealistic: the timing

22. We match the degree of income inequality in the data, but fall short of matching the extreme degree of wealth inequality observed in the real world. We discuss the standard and well-known reasons why this is so below.

of retirement is, for the most part, known. To deal with this unrealistic feature, we assume that there are perfect annuity markets for the retirees (neutralizing the influence of the risk of death on their behavior), and that workers' preferences have a certainty equivalence property (such that the *risk* of retirement does not affect workers' behavior in equilibrium).[23] These two assumptions are both realistic and convenient, in that they allow for the derivation of the aggregate consumption function, as we illustrate momentarily.

Specifically, we assume that agents have recursive Epstein–Zin preferences, which are defined as follows:

$$V_t^z = \left[(C_t)^\rho + \beta^z \mathbb{E}_t\{V_{t+1}|z\}^\rho\right]^{1/\rho}, \tag{7}$$

where C_t denotes consumption, V_t^z and β^z stand for the agent's $z \in \{w, r\}$ value function and the discount factor respectively, and $\sigma = \dfrac{1}{1-\rho}$ is the intertemporal elasticity of substitution.

Retirees and workers differ in two crucial respects. First, they have different discount factors. Because of the positive probability of death facing any retiree, his or her discount factor is the time preference parameter β multiplied by the probability of surviving into the next period:

$$\beta^w = \beta \tag{8}$$

and

$$\beta^r = \beta \cdot \gamma. \tag{9}$$

Second, the expectation of the value function next period differs between a worker and a retiree. In particular, a worker takes into account the possibility of retiring, so that his or her expectation of the value function in the next period is a probability-weighted sum of the values in the two states:

$$E_t\{V_{t+1}|w\} = \omega V_{t+1}^w + (1-\omega)V_{t+1}^r, \tag{10}$$

23. In particular, workers are assumed to have recursive Epstein and Zin (1991) preferences that generate certainty equivalent decision rules in the presence of income risk.

while the expectation of the value function of a retiree is simply given by

$$E_t\{V_{t+1}|r\} = V^r_{t+1}. \tag{11}$$

We now outline the problems of the two types of agents.

RETIREES Retirees consume out of saving and Social Security payments. Each period, some retirees die. We make the assumption—which is standard in the literature—that those who survive receive the proportional share of the proceeds. This means that the effective return faced by individual retirees is R_t/γ, higher than the ongoing interest rate R_t.[24]

Because the probability of death is independent of age and the government does not discriminate across retirees in its Social Security transfer policy, each retiree (irrespective of age) solves an identical problem, which is:

$$V^r_t = \max_{C^r_t}\left[(C^r_t)^\rho + \beta\gamma\mathbb{E}_t\{V^r_{t+1}\}^\rho\right]^{1/\rho}, \tag{12}$$

subject to the flow budget constraint

$$A^r_{t+1} = (R_t/\gamma)A^r_t - C^r_t + E^r_t, \tag{13}$$

where A^r_t stands for a retiree's assets, C^r_t are his or her consumption expenditures, and E^r_t is the Social Security and health care cost transfer.[25]

WORKERS Individuals are born workers and have no assets at the start of life. They consume out of asset wealth and their labor income net of taxes. Because of the demographic structure (in particular, the assumption that the probability of retirement is independent of age[26]), a worker's problem is effectively the same no matter the age. Each worker solves:

$$V^w_t = \max_{C^w_t}\left\{(C^w_t)^\rho + \beta\left[\omega V^w_{t+1} + (1-\omega)V^r_{t+1}\right]^\rho\right\}^{1/\rho} \tag{14}$$

subject to

24. For retirees *as a group*, wealth accumulates at the interest rate R_t, as the higher individual return cancels out with some retirees dying.

25. Our modeling of health care provision is very simple—we treat old-age health care cost as a lump-sum transfer, subsumed in the variable E.

26. Of course this is an unrealistic assumption. But as explained above, the effect of this assumption on workers' behavior is neutralized through the structure of preferences that exhibit a certainty equivalence property. The role of this assumption is thus only to simplify the model and achieve aggregation, with little cost to the economics.

$$A^w_{t+1} = R_t A^w_t + W_t - T_t - C^w_t, \tag{15}$$

where T_t are lump-sum taxes levied by the government.[27]

FIRMS The supply side of the model is extremely simple. Market are competitive. Production is carried out by firms employing capital and labor. The aggregate production function is

$$Y_t = K_t^{\alpha}(X_t N_t)^{1-\alpha}, \tag{16}$$

where N_t is the number of workers in the economy. There is exogenous technological progress and population growth—that is, $X_{t+1} = (1 + x)X_t$ and $N_{t+1} = (1 + n)N_t$. Perfect competition in factor markets means that the wage and the rental rate are equated to the marginal products of the factors: $W_t = \alpha \frac{Y_t}{N_t}$ and $R_t = (1 - \alpha)\frac{Y_t}{K_t} + (1 - \delta)$. Capital evolves according to the standard law of motion: $K_{t+1} = Y_t - C_t - G_t + (1 - \delta) K_t$.

GOVERNMENT The government consumes G_t each period, and pays retirees a total of E_t in Social Security and health care benefits. To finance its expenditures, the government levies a lump sum tax T_t on the workers. It can also issue one-period government bonds B_{t+1}. The government flow budget constraint is

$$B_{t+1} + T_t = R_t B_t + G_t + E_t. \tag{17}$$

Iterating forward gives the intertemporal budget constraint of the government:

$$R_t B_t = \sum_{v=0}^{\infty} \frac{T_{t+v}}{\prod_{z=1}^{v} R_{t+z}} - \sum_{v=0}^{\infty} \frac{G_{t+v}}{\prod_{z=1}^{v} R_{t+z}} - \sum_{v=0}^{\infty} \frac{E_{t+v}}{\prod_{z=1}^{v} R_{t+z}}. \tag{18}$$

27. There are two key channels through which life-cycle considerations affect workers' behavior. First, a worker takes into account the fact that, with probability $1 - \omega$, he or she becomes a retiree. This means that, relative to the representative agent case, he or she discounts the future stream of wages by more; effectively, this is the *saving-for-retirement* effect. Mechanically, a larger discount rate reduces the value of human wealth in the consumption function, thus leading to lower consumption and higher saving. Second, a worker discounts the future stream of wealth more because he or she anticipates that inevitably there will come a time when he or she becomes a retiree, facing the sad truth that life is finite. With finite life, wealth can be smoothed out across fewer periods, so its marginal utility value is lower. This effect shows up as a higher effective discount rate applied to future wealth.

That is, the difference between the present discounted value of government revenue and spending must be exactly equal to the current value of the outstanding debt.

Government policy is exogenous. In particular, it is characterized by the four ratios, $\bar{g}_t$, $\bar{b}_t$, $\bar{e}_t$, $\bar{h}_t$ of, respectively, government consumption, debt, Social Security, and health care spending to GDP:

$$G_t = \bar{g}_t Y_t \tag{19}$$

$$B_t = \bar{b}_t Y_t \tag{20}$$

$$E_t = \left(\bar{e}_t + \bar{h}_t\right) Y_t. \tag{21}$$

Given the paths of G_t, E_t, and B_t, taxes adjust to satisfy the intertemporal budget constraint.

EQUILIBRIUM In this economy, markets are competitive and agents take prices as given. Formally, a competitive equilibrium is a sequence of quantities and prices such that (1) households maximize utility subject to their budget constraints; (2) firms maximize profits subject to their technology constraints; (3) the government chooses a path for taxes, compatible with intertemporal solvency, to finance debt, spending, and transfers; and (4) all markets clear.

Online appendix C contains the details of the derivation of the equilibrium conditions of the model. The individual policy functions within the two groups—workers and retirees—aggregate up nicely. Aggregating the two consumption levels, we derive the aggregate consumption function:

$$C_t = C_t^w + C_t^r = \pi_t \{(1 - \lambda_t) R_t A_t + H_t + S_t^w + \epsilon_t (\lambda_t R_t A_t + S_t^r)\}. \tag{22}$$

In this consumption function, π_t denotes each worker's MPC out of wealth, and $\pi_t \epsilon_t$ is the MPC of each retiree. These MPCs multiply the total wealth of each group of consumers—with a slight abuse of notation, A_t now denotes aggregate financial wealth, H_t is aggregate human wealth (the net present value of future wages), and S_t stands for the aggregate value of Social Security and health care payments. Compared with a standard model, the only additional state variable is the share of wealth held by retirees, λ_t, which fully captures the heterogeneity in the economy.

Table 3. Calibration of the Model

Parameter	*Description*	*Calibration*
Preferences and technology		
β	Discount factor	0.98
σ	Intertemporal elasticity of substitution	0.5
α	Capital share	0.33
δ	Depreciation rate	0.1
x	Rate of technological change	1.51%
Demographics		
n	Gross population growth rate	1.35%
$\frac{1}{1-\omega}$	Average length of working life (years)	47.6
$\frac{1}{1-\gamma}$	Average length of retirement (years)	10.5
Government ratios		
$\bar{b}$	Government debt to GDP	0.18
$\bar{g}$	Government consumption to GDP	0.14
$\bar{e}$	Social Security spending to GDP	0.04
$\bar{h}$	Old-age health care spending to GDP	0.02

Source: Authors' calculations.

The total supply of assets is the sum of capital stock K_t and government debt B_t, so that the equilibrium requires

$$A_t = A_t^w + A_t^r = K_t + B_t, \tag{23}$$

that is, households asset demand equals the asset supply.

CALIBRATION AND THE INITIAL STEADY STATE OF THE LIFE-CYCLE MODEL Despite the richness of the economics, the model is parsimonious and relatively straightforward to calibrate. We set the preferences and technology parameters at the standard values in the macroeconomic literature (table 3). The growth rate of technological change, the demographics parameters, and the government policy ratios are all calibrated to match the data in the advanced economies in 1970.

Because there are population growth and technological progress in this economy, the steady state equilibrium takes the form of a balanced growth path, where all variables grow at a constant gross rate equal to $(1 + n)(1 + x)$. We can characterize the equilibrium by expressing all variables as ratios in units of effective labor (defining, for any variable $Z_t, z_t \equiv \frac{Z_t}{X_t N_t}$).

Table 4. The 1970s Steady State[a]

Variable	*Description*	*Value*
ψ	Ratio of retirees to workers	0.19
R	Real gross interest rate	1.045
ϵ	Ratio of retirees' to workers' MPCs	2.01
π_w	Workers' MPC	0.06
π_r	Retirees' MPC	0.13
λ	Share of retirees' wealth in total wealth	0.17
y	Output	1.50
Ratios (in proportion to output)		
c	Consumption	0.57
c_r	Consumption of retirees	0.11
c_w	Consumption of workers	0.45
a	Assets	2.42
a_r	Assets of retirees	0.40
a_w	Assets of workers	2.03
h	Human capital	4.23
i	Investment	0.27
k	Capital	2.25
τ	Taxes	0.21
s	Social Security wealth of the retirees	0.50
s_w	Social Security wealth of the workers	0.91

Source: Authors' calculations.
a. MPC = marginal propensity to consume.

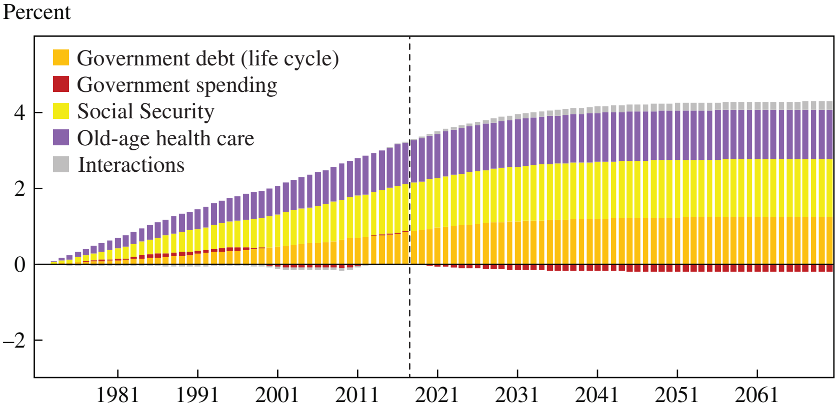
Table 4 shows the key variables along the initial (early-1970s) balanced growth path. The interest rate is 4.5 percent.[28] As we pointed out above, the key feature of this economy is the heterogeneity in MPCs between workers and retirees. Indeed, the endogenous MPC of retirees is over twice that of the workers. The additional state variable λ—the ratio of retirees' wealth in total wealth—takes a plausible value of 17 percent. Ratios of aggregate consumption, investment, capital, and assets to output also match the stylized facts from the data well.

THE SIMULATION EXERCISE We now explore how the model economy reacts to changes in government policy. We study four policy levers: government debt, government spending, old-age Social Security, and health care transfers.

We carry out the following experiment. Starting the economy in the initial 1970s steady state, we feed the model with the policy profiles depicted in figure 6. Once announced, the profile of these shifts is fully

28. With a growth rate of 2.9 percent a year, the economy is dynamically efficient.

Figure 9. The Simulated Impact of Government Policies on the Equilibrium Real Interest Rate in the Life-Cycle Model, 1971–2071[a]



Source: Authors' calculations.

a. This figure shows how the equilibrium real interest rate adjusts to the exogenously given paths of government debt, spending, and old-age Social Security transfers depicted in figure 6. The 2017 values given in figure 6 are assumed to be the new steady state values.

anticipated by the agents. Beyond the current date, we assume that future policy ratios remain constant at their 2017 values.[29] We then compute the transition path toward this new steady state.

Our focus is on the response of the interest rate to these policy shifts. Figure 9 contains the main result of this section: the total response of the interest rate to the policy changes discussed above. This response is quantitatively large; according to the model, government policies have pushed up on the equilibrium interest rate by about 3.2 percentage points over the past 50 years. Moreover, the model suggests that further upward pressure is to be expected as the economy settles at the new steady state. All the policies except government spending—which did not change much—play an important role. The final set of bars in figure 9, labeled "interactions," is the additional effect on the interest rate from the (nonlinear) synergies between the three different policies.[30]

29. This is a conservative assumption, as one may reasonably expect the upward drift in both debt and Social Security spending to continue, at least for some time.

30. More precisely, the interaction effect exists because the final steady state is a nonlinear system of equations. These nonlinearities make the overall effect of several exogenous changes different, in general, from the sum of the parts.

IV.C. A Model of Precautionary Saving

We now turn to the model of precautionary behavior, which is a continuous-time version of the Aiyagari and McGrattan (1998) economy. The population consists of a large number of infinitely lived individuals of measure 1. Every individual is ex-ante identical, but people face shocks to their income that they cannot fully insure against: markets are incomplete. As a result of this idiosyncratic risk, individuals experience different income histories and thus accumulate different levels of wealth. All the risk is at the individual level; for simplicity, we abstract from aggregate uncertainty.

Our goal here is to quantitatively assess the influence government debt has on precautionary behavior. In other words, how different is the prevailing interest rate when the government debt-to-GDP ratio is 18 percent versus when it is 68 percent?[31]

A BRIEF OUTLINE OF THE MODEL An individual chooses consumption and asset holdings to maximize his or her expected utility, subject to the flow budget constraint, the consumption nonnegativity constraint, the borrowing constraint, and a realization of the idiosyncratic income shock:

$$\max_{\{c_t\}_{t>0}} \mathbb{E}_0 \int_0^\infty e^{-\rho t} \frac{c_t^{1-\sigma}}{1-\sigma} dt \tag{24}$$

subject to

$$\dot{a}_t = (1-\tau) w_t e_t + (1-\tau) r_t a_t - c_t$$

$$c_t \geq 0$$

$$a_t \geq \underline{a}$$

$$e_t \in \{z_1, \ldots, z_n\},$$

31. Our model is highly stylized and abstracts from important features present in more advanced and larger models in the literature. We view our model here as an early attempt to quantify the precautionary saving channel of government debt. Richer features may usefully be incorporated in future attempts to answer this question. For analysis of saving rates across the distribution, see Straub (2017) and Fagereng and others (2019). For evidence on the differential rates of return, see Fagereng and others (2016). For models with multiple assets or a more careful analysis of the constraint—both of which contribute to a better match to the empirical distribution around the borrowing constraint, see Kaplan, Violante, and Weidner (2014); Kaplan, Moll, and Violante (2018); and Achdou and others (2017). For the state-of-the-art calibration of the income process, see Guvenen and others (2015). We conjecture that a richer model with some of the above features would likely predict larger effects of government policy.

where c_t is individual consumption, a_t are individual asset holdings (and $\dot{a}_t$ denotes the time derivative, that is, saving), r_t is the real (net) interest rate, w_t is the wage, and e_t is the idiosyncratic shock to a household's productivity. The household cannot insure against this idiosyncratic uncertainty. The government levies a proportional tax rate τ on both labor and capital income.[32]

The supply side is identical to that in the previous model: the production function is Cobb–Douglas, and there is perfect competition in all markets. The government issues bonds and collects taxes to finance its consumption and transfers. The government budget constraint is

$$\dot{B}_t = G_t + r_t B_t - \tau(w_t + rA_t), \tag{25}$$

which says that the change in government debt is equal to the government funding gap: government consumption G_t plus interest payments $r_t B_t$ minus the tax revenue.

Online appendix D presents the definition and solution of the equilibrium of this economy.

PARAMETERIZATION We choose the values of the parameters in the precautionary saving model to match the typical values in the literature and to be broadly consistent with the life-cycle model given above (table 3). We set the capital share at 0.33, the rate of time preference at 0.04, the depreciation rate at 10 percent, and the intertemporal elasticity of substitution at 0.50.

We next calibrate the income process. Intuitively, the size and persistence of income shocks will determine the strength of the precautionary saving motive, the degree of inequality, and the proportion of households close to or at the borrowing constraint. These outcomes will in turn determine the potency of government financing policy. In the real world, individual income varies over time for a host of reasons. We do not model these causes here. Instead, we make sure that the income process in our model reflects these uncertainties. Specifically, we follow Ana Castañeda, Javier Díaz-Giménez, and José-Víctor Ríos-Rull (2003) and Christopher Winter (2016), and we thus calibrate the income process

32. The assumption of a proportional tax rate is natural in a model with income and wealth heterogeneity. With lump-sum taxation, the poorest households would find themselves unable to pay the tax bill. Note that even though the tax is proportional it does not distort the labor supply decisions because the labor supply is inelastic.

to match aggregate income inequality in the OECD. There are four productivity and income states:

$$e \in \{0.20, 0.55, 0.80, 5.43\}. \tag{26}$$

The corresponding matrix of Poisson intensities is

$$P = \begin{pmatrix} 0.07^{-} & 0.04 & 0.02 & 0.001 \\ 0.03 & 0.13^{-} & 0.01 & 0.001 \\ 0.001 & 0.08 & 0.09^{-} & 0.011 \\ 0.1 & 0.02 & 0.06 & 0.17^{-} \end{pmatrix}, \tag{27}$$

where the values on the main diagonal marked with superscript $^{-}$ indicate the intensity of leaving the current state.

Given this income process, the distributional outcomes in the equilibrium of our model are broadly in line with those observed in the data: the income Gini coefficient is 0.32, close to the OECD average, and the income process is highly persistent.[33]

RESULTS We now compare the two stationary equilibria of the model, one with the government debt-GDP ratio set at 18 percent, and another at 68 percent, to see what the impact of such a higher pool of assets is on the interest rate. Because a larger amount of assets allows households to better insure against individual uncertainty, we expect the interest rate to be higher when government debt is high. The simulation results confirm this intuition: the increase in the public debt-GDP ratio observed in the data implies a real interest rate that is 66 basis points higher in equilibrium (table 5). Though not insignificant, such an increase is smaller than the other channels we identified above.

IV.D. Summary and Discussion

To summarize this section, our analysis underscores the importance of secular public policy shifts in accounting for changes in the equilibrium interest rate. The natural corollary of our findings is that government

33. Castañeda, Díaz-Giménez, and Ríos-Rull (2003) compare the across-the-income-distribution mobility statistics implied by their model with those observed in the data and conclude that the simple model does reasonably well in capturing the persistence moments.

Table 5. Equilibria in the Precautionary Saving Model

Aspect of model	*Low-debt equilibrium*	*High-debt equilibrium*
Government debt to GDP	0.18	0.68
Government consumption to GDP	0.14	0.14
Average tax rate	0.35	0.36
Real interest rate	4.50	5.16
Private capital to GDP	2.56	2.40
Income Gini coefficient	0.32	0.32
Fraction of individuals at the constraint	0.09	0.09

Source: Authors' calculations.

intra- and intertemporal transfer policy is, in principle, an effective tool that can affect equilibrium interest rates in the economy. Similar policy implications have been discussed previously by Narayana Kocherlakota (2015) and Caballero and Farhi (2018).

One objection to our analysis might be that economic agents—consumers, investors, firms, and the like—may in fact be more Ricardian than we currently assume. Our response to this is threefold. First, in section II we presented a broad range of empirical evidence that is inconsistent with the Ricardian Equivalence proposition. Second, in our framework, Ricardian Equivalence does not hold despite fully rational expectations and no information asymmetries; indeed, it would be irrational to be Ricardian in the economy we describe. Third, and relatedly, the assumptions that lead to rejection of Ricardian Equivalence are rather natural—first, people retire; second, people die; third, some people are credit constrained; and fourth, some people face risks they find hard to insure. All these considerations make us comfortable with our assumptions that the Ricardian offset is imperfect.

At this point, it is also useful to highlight that wide uncertainty bands surround our point estimates, including those coming out of the models discussed above. Like all theory models, these tools are built upon a set of uncertain assumptions, and as such are only rough approximations of reality—this is especially true for models as minimalistic and transparent as ours. Even abstracting from model misspecification, there is a wide range of plausible parameter values with which to calibrate these models. A different combination of parameters will produce quantitatively different results. We come back to the robustness of our analysis in online appendix E. Having said that, the combination of a range of empirical studies together with directional guidance from the theory suggest that there are strong reasons to conclude that the government policies we have

Table 6. Demographic Transition in the Advanced Economies, 1970–2030

Year	*Growth of 20+ population*	*Retirement age*	*Years working*	*Years in retirement*
1970	1.4	67.6	47.6	10.5
1975	1.3	66.6	46.6	12.3
1980	1.2	66.1	46.1	13.4
1985	1.1	65.1	45.1	15.0
1990	0.9	64.7	44.7	16.1
1995	0.8	63.8	43.8	17.5
2000	0.7	63.6	43.6	18.6
2005	0.8	64.1	44.1	18.9
2010	0.7	64.8	44.8	18.8
2015	0.4	65.5	45.5	18.7
Projection:				
2020	0.2	66.1	46.1	18.6
2025	0.2	66.8	46.8	18.4
2030	0.2	67.5	47.5	18.3

Sources: United Nations data; OECD data.

scrutinized here have put significant upward pressure on the safe neutral real rate over the past several decades.

V. Validating the Models by Assessing the Underlying Weakness in R^*

Our simulation analysis concluded that the major shifts in governments' policies over the past 50 years facilitated a significant transfer of resources from low-MPC to high-MPC individuals and allowed households to better self-insure against idiosyncratic shocks. All else being equal, added together these shifts would have pushed interest rates in the advanced world up by about 3.6 percentage points. But of course all else was not equal. In this section, we validate our models by showing that, when used to assess the impact of some of the private sector forces that have been highlighted by the literature, these models produce the quantitative effects that are plausible and in line with the existing findings. Specifically, our framework can readily be used to quantify the impact of the demographic transition, the decline in expected trend productivity growth, and the rise in income inequality on the long-term interest rate.

Table 6 documents the major demographic transition that has been under way in the advanced economies for the past 50 years. Population growth in the developed economies has fallen rapidly in past decades, from about 1.4 percent a year in the 1970s to less than 0.4 percent today. This trend is expected to continue; in fact, the latest UN projections suggest that population in the advanced economies will start shrinking in about 2050. As population growth has decelerated, life expectancy has gone up significantly, and retirement ages have not kept up. As a result, the average length of retirement is nearly twice what it was in the 1970s. This positive development carries significant implications for life-cycle budgeting and thus for the balance of desired saving and investment.

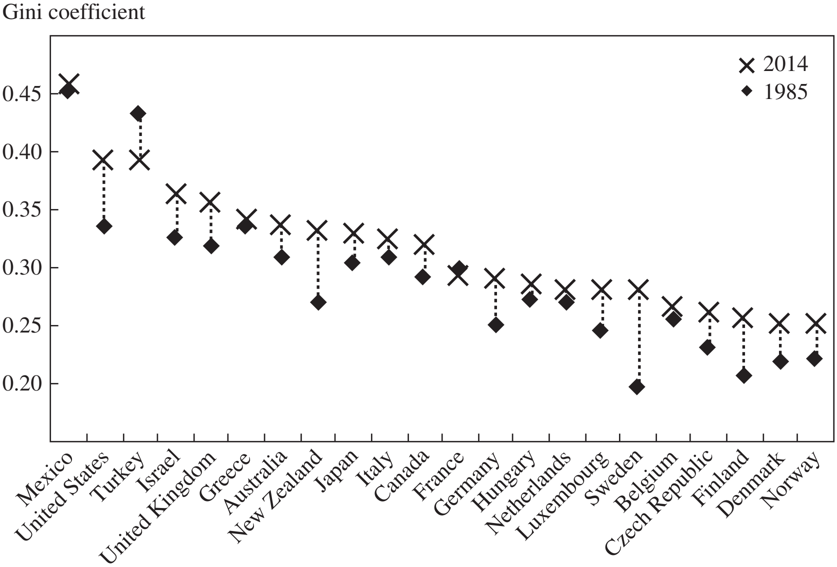
The slowdown in the pace of expected long-run growth has similar implications. Our modeling framework inherits the property shared by essentially all dynamic macroeconomic models, namely, that the long-run equilibrium interest rate is linked to expected future consumption growth. This relationship—the Euler Equation, or the dynamic investment–saving curve—is the result of the intertemporal optimization of households, which choose how much to consume today versus tomorrow (hence determining the growth rate of their consumption) based on the interest rate. In general equilibrium, the expectations of future consumption growth in the long run coincide with the expectations of total factor productivity (TFP) growth. Hence the theory suggests that real rates and expected productivity growth ought to be linked.[34]

This prediction of the theory is, however, more tenuous in practice. In an early contribution to this topic, Carroll and Summers (1991) established that, across countries, consumption growth and income growth are tightly linked and follow each other, and that households with more steeply rising income profiles tend to save more, not less. These findings—which are inconsistent with the standard permanent-income hypothesis and the life-cycle model—have been rationalized in the literature with buffer-stock models of saving (whereby households face uncertain income process, similarly to our second model) and introducing consumption habits in household preferences.[35] Although our models attenuate the link between

34. In a representative agent, infinite-horizon economy, the Euler Equation takes a particularly straightforward form, whereby the long-run consumption growth rate and the interest rate are linked linearly, with the coefficient equal to the intertemporal elasticity of substitution. Within our framework, that link is still there, although it is attenuated by finite horizons and borrowing constraints: intuitively, the interest rate is relatively "less important" in driving consumption growth, as other factors (such as the possibility of death or credit constraints) come into play. This implies that a given change in the expectations of future consumption growth—say, driven by news about TFP—will require a *larger* response of the interest rate to restore equilibrium.

35. See Deaton (1991); Carroll (1997); and Carroll, Overland, and Weil (2000)—and the literature that followed.

Figure 10. The Gini Coefficient of Disposable Household Income across the OECD, 1985 and 2014[a]



Source: OECD Database on Household Income Distribution and Poverty.
a. This figure shows the Gini coefficient of disposable income, adjusted for household size.

interest rates and consumption choices in line with these findings, nonetheless we urge a significant degree of caution when interpreting the results on the link between TFP and R^*. Our preferred interpretation is that the low interest rates today are chiefly a symptom of a demand-side problem. We return to this issue in the final section of the paper, where we discuss policy implications.

These caveats notwithstanding, numerous studies—for instance, that by Gustavo Adler and others (2017)—reach the conclusion that trend growth rates of both productivity and of TFP have declined significantly in the advanced economies—first in the early 1980s, when TFP growth halved from about 2 to 1 percent a year; and then again in the mid-2000s. Also, the macroeconomic models we use do suggest that such deterioration should have dragged on neutral real interest rates.

The third trend we quantify is the rise in income inequality, which has increased in the United States and many other advanced economies (figure 10). Our second model is well suited to give us an estimate of this shift on the real rate of interest. To trace out the effects of rising inequality in this model, we recalibrate the income process in such a way as to match the increase in income Gini coefficient in the OECD since the 1970s. Our calculations implicitly assume that ex-post inequality is driven by larger variance of individual income shocks, which constitutes a source of additional uncertainty for individual workers. An alternative view is that the increase in inequality is a consequence of shifts being more tightly linked to heterogeneity across households that is known ex-ante. The distinction is important because only the former kind of shift would lead to an increase in precautionary behavior. Because it is predictable, the latter shift is not associated with heightened risk. There is a long-standing debate about the merits of the two formulations in the literature.[36] The recent work by Fatih Guvenen and others (2015) has established the large departures of log-normality in the individual income changes; in particular, earnings changes display strong negative skewness and extremely high kurtosis. Important for our interpretation is their finding that large shocks at the top of the income distribution tend to be very persistent. We view these results as supportive of the gist of our exercise, which interprets the increased disparity between the poor and the rich as going hand in hand with an increase in ex-ante uncertainty. Given the lack of a clear consensus in the literature, it is possible that we overestimate the impact of inequality on real rates in this exercise. In any case, there likely are other powerful ways in which higher inequality has acted to depress rates, which we miss from our framework (and which we discuss momentarily).

To validate our models and to explore the implications of these trends for the equilibrium real interest rate, we perform this exercise: In the life-cycle model, we calibrate the changes in demographic transition probabilities, ω and λ, to match the trends depicted in the final two columns of table 6. We then feed in the series for population and TFP growth rates to match the evidence in the first column of table 6 and as given by Adler and others (2017). We use the United Nations' demographic projections to inform the path of demographics out to 2050, and we assume that the terminal 2050 values are the steady state. We do not have a strong prior as to the path for future TFP growth, and we are well aware of the wide range of existing and plausible views. Aiming for a scenario that reflects the mode of these expectations, we assume that the TFP growth rate picks up from around

36. Classic references include Lillard and Weiss (1979), MaCurdy (1982), and Guvenen (2009).

zero in the latest available data to 0.7 percent in the long run.[37] This pickup in TFP growth is broadly in line with the Congressional Budget Office's assumption for the pickup of TFP growth in the United States (CBO 2019).

In the precautionary saving model, we recalibrate the income process,[38] and we compare the steady states of the economy under the two calibrations.[39]

To reiterate, within each of the two models, we feed in the (model-specific) set of shocks all at the same time, thereby providing—within each model—an internally consistent laboratory to study this wide range of heterogeneous trends. What we miss are the potential interactions across the two models. We assume that the comparable calibration across the two frameworks makes the results comparable, and that simply adding the estimates of the impact on R^* over the transition across the models results in a consistent picture. But ultimately, only the framework for analysis of all the forces that we consider—and perhaps further ones—in a single unifying setting would provide a definitive answer to these doubts. This avenue of inquiry is left for future research.

Table 7 and figure 11 summarize the key results of this exercise. First, in section II we estimated that the neutral real rate declined by over 3 percentage points between 1970 and 2017. In sections III and IV, we argued that public policies have pushed rates up. Our models suggest that, together, the policies we have considered have pushed rates up by nearly 4 percentage points to date. This suggests that the private sector R^* may have declined by about 7 percentage points. The private sector forces we consider add up to a drag of 5.5 percentage points, leaving over 1 percentage point of the decline in private sector R^* unaccounted for. These results are in line with previous papers that have attempted the quantification of the different forces at play (Eggertsson, Mehrotra, and Robbins 2019; Carvalho, Ferrero, and Nechio 2016; Gagnon, Johannsen, and Lopez-Salido 2016). This makes us confident that the large quantitative effect of government policies

37. There is very large uncertainty around any long-term forecast of the TFP growth rate. In particular, research has shown that current-decade growth of productivity holds little information as to the growth in the following decade. Perhaps naturally, the commentators are split on the prospects for innovation and productivity. See, for example, Brynjolfsson and McAfee (2014) and Gordon (2016) for two perspectives from the opposite ends of a spectrum.

38. In particular, we change the income received in the highest income state. This is motivated by the fact that the increase in income inequality has been concentrated at the very top of the distribution, as documented by Piketty (2014) and others.

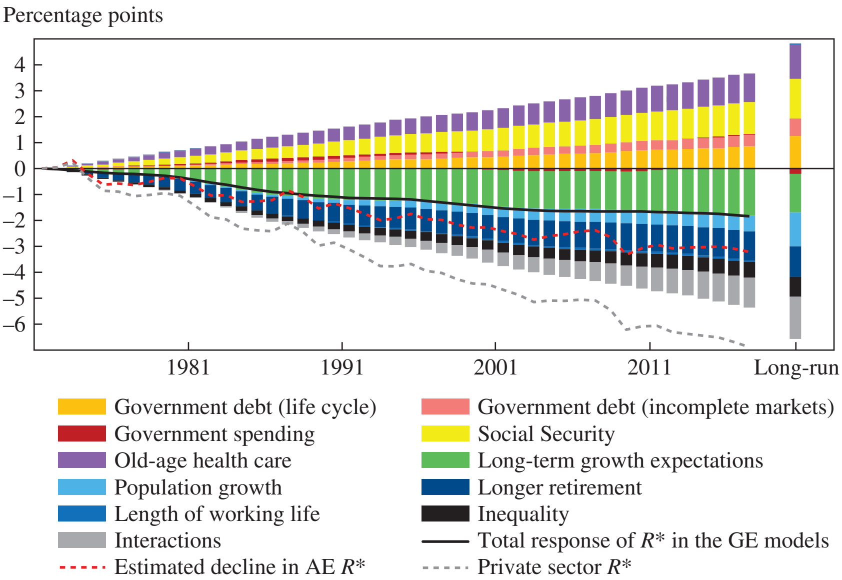
39. We obtain the dynamic path by assuming that the effect builds steadily, including over the next decade. Our treatment of the dynamics is thus crude. We leave the analysis of the dynamic adjustment path for future work.

Table 7. Decomposition of the Decline in the Neutral Real Interest Rate in the Advanced Economies[a]

Aspect of decomposition	*1970–2008*	*1970–2017*	*1970–2070*
Estimated decline in AE R^* (section II)	–2.7	–3.2	
Public policies			
Government debt (life cycle)	0.6	0.8	1.2
Government debt (incomplete markets)	0.3	0.4	0.7
Government spending	–0.1	0.0	–0.2
Social Security	1.0	1.2	1.5
Old-age health care	0.9	1.1	1.3
Total impact of public policies	*2.8*	*3.6*	*4.5*
Implied decline in private sector R^*	–5.4	–6.9	
Selected private sector forces			
TFP growth	–1.5	–1.8	–1.5
Population growth	–0.5	–0.6	–1.3
Longer retirement	–1.0	–1.1	–1.2
Length of working life	–0.1	–0.1	0.0
Inequality	–0.6	–0.7	–0.9
Interactions	–0.8	–1.1	–1.6
Total private sector forces	–4.4	–5.5	–6.5

Source: Authors' calculations.
a. TFP = total factor productivity. All values are in percentage points.

Figure 11. Changes in the Equilibrium Real Interest Rate as a Result of Policy, Demographic, and Technological Shifts, 1971–2022



Source: Authors' calculations.

that we estimated are credible and not just a result of model-specific assumptions or calibration.

Unsurprisingly, the forces that we consider in this exercise cannot account for the full extent of the decline in equilibrium rates, with over 1 percentage point left unexplained in our preferred calibration. Our models miss some of the secular forces that likely pushed neutral rates lower over the past 40 years. One omission is the increasing concentration and the associated increase in market power of firms in the United States and other advanced countries (Farhi and Gourio 2018). Another force is driven by the finding that propensities to save are higher for those with high permanent income (Carroll 2000; Dynan, Skinner, and Zeldes 2004). In light of these findings, our simulations likely understate the full impact of the increase in permanent income inequality. Using a model that captures this mechanism, Ludwig Straub (2017) estimates that the rise in inequality may have pushed down on the real equilibrium interest rate in the U.S. by about 1 percentage point through this channel. The decline in the price of capital goods may have contributed to lower investment propensities, further decreasing the neutral real rate (Sajedi and Thwaites 2016). Finally, changes in the tax code—particularly the decline in overall tax progressivity in some jurisdictions—may have been a public sector force that depressed interest rates. We leave more detailed investigation of these forces for future research.

VI. Conclusion

We draw three main conclusions from the analysis in this paper. First, the neutral real rate for the industrial world has trended downward for the last generation, and this is best understood in terms of changes in private sector saving and investment propensities. In the face of neutral real rate estimates, past trends in indexed bond yields, and measures of real swap yields, this conclusion seems inescapable. It is also noteworthy that current real rates appear to be quite well predicted by prefinancial crisis trends. We believe that these trends are best analyzed in terms of changes in saving and investment propensities or equivalently in terms of trends in desired wealth holdings by consumers and desired capital accumulation by producers. Although factors involving liquidity, scarcity, and risk no doubt bear on levels of real interest rates, we find it highly implausible that they are the main factor accounting for the trend movements. The movements are too large and too pervasive across assets and the fluctuations in spreads are too small and lacking in the trend for these factors to account for the observed trends in the data.

Second, the neutral real rate would have declined substantially more over the last generation but for increases in government debt and expansions in social insurance programs. Both straightforward extrapolations of existing rules of thumb regarding debt and deficit effects on interest rates and calculations using workhorse general equilibrium models suggest that fiscal policies have operated to raise real interest rates by several hundred basis points over the last generation. Though this conclusion is dependent on our rejection of Ricardian Equivalence, we see nothing that leads us to believe that increased government debt automatically calls for increased saving or that pay-as-you-go Social Security programs alter bequests for most families. The specific magnitudes are very uncertain, but open economy aspects and the possibility suggested by our analysis—that budget deficits emerge in response to excesses of private saving over private investment—lead us to think that we are more likely to understate than overstate the extent of fiscal support for real interest rates in recent years.

Third, the implication of our analysis that but for major increases in deficits, debt, and social insurance neutral real rates in the industrial world would be significantly negative by as much as several hundred basis points suggests substantial grounds for concern over secular stagnation. From the perspective of our analysis, the private economy is prone to being caught in an underemployment equilibrium if real interest rates cannot fall far below zero. Where full employment has been achieved in recent years, it has either been through large budget deficits, as in the United States or Japan, or through large trade surpluses, as in Germany. It is worth considering that in the United States during the period before the financial crisis, negative real short-term interest rates, a huge housing bubble, erosion of credit standards, and expansionary fiscal policy were only sufficient to achieve moderate growth. Adequate growth in Europe was only maintained through what in retrospect appears to have been clearly unsustainable lending to the countries on the so-called periphery.

What does our analysis say about stabilization policy? Most obviously, it says that traditional levels of interest rates combined with balanced budgets or even stable debt-GDP ratios are a prescription for recession. If policymakers wish to avoid output being demand constrained, they must do some combination of accepting high and rising deficits and government debt levels, living with real interest rates very close to zero or negative, and finding structural policies that promote investment or reduce saving.

Blanchard (2019) makes the argument that traditional views about fiscal policy likely reflect excessive concern about debt when real interest rates are very low and are likely to remain low for a long time to come. The sustainability of a given level of deficit or debt is greater when interest rates are low than when they are high. Nonetheless, it has to be acknowledged that the U.S. economy appears to be slowing to below potential growth despite projected primary deficits that will lead even on very favorable interest rate assumptions to steadily growing debt-to-GDP ratios that will ultimately set historical records. There is no guarantee that deficits sufficient to maintain positive neutral real rates will be associated with sustainable debt trajectories. Indeed, the Japanese experience suggests that this may not be the case.

Another possibility is the use of monetary policies that induce significantly negative real rates. This might be achieved through setting negative nominal rates, raising or adjusting inflation targets (for example, through targeting the average rate of inflation and thus "making up" for the past errors), or using unconventional monetary policies such as quantitative easing to achieve the equivalent of reductions in real rates. These approaches raise three issues. First, given that historically rates have been reduced by 500 basis points or more to mitigate recessions in industrial countries, there is the question of whether enough room can be generated to stabilize the economy when the next downturn hits. Second, there are questions about whether, starting at very low rates, further rate reductions are actually stimulative. Eggertsson and others (2019) suggest that negative nominal rates actually may interfere with financial intermediation. Third, there is a range of concerns about the possible toxic effects of low rates—including suggestions that they make bubbles and overleveraging more likely as they encourage risk taking, and that they may lead to a misallocation of capital by reducing loan payment levels and required rates of return, reinforcing monopoly power, benefiting the old at the expense of the young, and making the funding of insurance and pension obligations more difficult.

A final possibility is structural measures that reduce saving or promote investment. Clearly, regulatory policies that encourage investment without sacrificing vital social objectives are desirable. The extent to which these are available is very much open to question. Investment incentives will also operate to raise demand. Policies that reduce the need for retirement saving, such as strengthening Social Security, or that improve social insurance, will increase aggregate demand even if operated on a balanced budget basis. So will policies that redistribute income from those with lower to those with higher propensities to consume.

It is tempting to suggest that any measure that increases productivity growth will operate to raise neutral real rates as consumers seek to spend more out of higher expected future incomes and firms increase their investment demand. Effects of this kind are indeed suggested by our formal model. We are not sure of their validity in practice. As Carroll and Summers (1991) point out, growth accelerations internationally have typically been associated with declining rather than increasing real rates, and there is not much evidence that consumers are that forward looking, especially if the reforms are associated with transitional costs and heightened short-term uncertainties. Moreover, in policy discussions, central bankers usually cite stronger productivity as an antidote to inflation and therefore as a reason not to raise rates. Short-term productivity gains that reduce costs and inflation may act to elevate realized interest rates above the neutral rate, further worsening the demand imbalance.

All this suggests that if secular stagnation is avoided in the years ahead, it will not be because it is somehow impossible in a free market economy, but instead because of policy choices. Our conclusions thus underscore the urgent priority for governments to find new, sustainable ways of promoting investment to absorb the large supply of private saving and to devise novel long-term strategies to rekindle private demand.

ACKNOWLEDGMENTS We are grateful to the editors, Janice Eberly and James Stock, and to our discussants, Arvind Krishnamurthy and Gauti Eggertsson, for their detailed comments that much improved the paper. We also thank Ricardo Reis, Daniele Siena, Anna Stansbury, and participants in the NuCamp conference in Oxford for insightful comments on the earlier versions. Łukasz Rachel thanks the Department of Economics at Harvard University, where much of this project was completed during his visit as a Fulbright Fellow, and the U.S.-U.K. Fulbright Commission for financial support.

References

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2017. "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach." NBER Working Paper 23732. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w23732.

Adler, Gustavo, Romain Duval, Davide Furceri, Sinem Kiliç Çelik, Ksenia Koloskova, and Marcos Poplawski-Ribeiro. 2017. "Gone with the Headwinds: Global Productivity." IMF Staff Discussion DN/17/04. Washington: International Monetary Fund. https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2017/04/03/Gone-with-the-Headwinds-Global-Productivity-44758.

Aiyagari, S. Rao. 1993. "Uninsured Idiosyncratic Risk and Aggregate Saving." *Quarterly Journal of Economics* 109, no. 3: 659–84.

Aiyagari, S. Rao, and Ellen R. McGrattan. 1998. "The Optimum Quantity of Debt." *Journal of Monetary Economics* 42, no. 3: 447–69.

Attanasio, Orazio P., and Agar Brugiavini. 2003. "Social Security and Households' Saving." *Quarterly Journal of Economics* 118, no. 3: 1075–1119.

Attanasio, Orazio P., and Susann Rohwedder. 2003. "Pension Wealth and Household Saving: Evidence from Pension Reforms in the United Kingdom." *American Economic Review* 93, no. 5: 1499–1521.

Barro, Robert J. 1974. "Are Government Bonds Net Wealth?" *Journal of Political Economy* 82, no. 6: 1095–1117.

Barro, Robert J., and Glenn M. MacDonald. 1979. "Social Security and Consumer Spending in an International Cross Section." *Journal of Public Economics* 11, no. 3: 275–89.

Blanchard, Olivier. 1985. "Debt, Deficits, and Finite Horizons." *Journal of Political Economy* 93, no. 2: 223–47.

———. 2019. "Public Debt and Low Interest Rates." *American Economic Review* 109, no. 4: 1197–1229.

Brook, Anne-Marie. 2003. "Recent and Prospective Trends in Real Long-Term Interest Rates." Working Paper 367. Paris: Organization for Economic Cooperation and Development, https://doi.org/10.1787/514820262776.

Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W. W. Norton.

Caballero, Ricardo J., and Emmanuel Farhi. 2018. "The Safety Trap." *Review of Economic Studies* 85, no. 1: 223–74.

Caballero, Ricardo J., Emmanuel Farhi, and Pierre-Olivier Gourinchas. 2016. "Safe Asset Scarcity and Aggregate Demand." *American Economic Review* 106, no. 5: 513–18.

Carroll, Christopher D. 1997. "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis." *Quarterly Journal of Economics* 112 , no. 1: 1–55.

———. 2000. "Why Do the Rich Save So Much?" In *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, edited by Joel B. Slemrod. Cambridge, Mass.: Harvard University Press.

Carroll, Christopher D., Jody Overland, and David N. Weil. 2000. "Saving and Growth with Habit Formation." *American Economic Review* 90, no. 3: 341–55.

Carroll, Christopher, Jiri Slacalek, Kiichi Tokuoka, and Matthew N. White. 2017. "The Distribution of Wealth and the Marginal Propensity to Consume." *Quantitative Economics* 8 , no. 3: 977–1020.

Carroll, Christopher D., and Lawrence H. Summers. 1991. "Consumption Growth Parallels Income Growth: Some New Evidence." In *National Saving and Economic Performance*, edited by B. Douglas Bernheim and John B. Shoven. Chicago: University of Chicago Press.

Carvalho, Carlos, Andrea Ferrero, and Fernanda Nechio. 2016. "Demographics and Real Interest Rates: Inspecting the Mechanism." *European Economic Review* 88: 208–26.

Castañeda, Ana, Javier Díaz-Giménez, and José-Víctor Ríos-Rull. 2003. "Accounting for the U.S. Earnings and Wealth Inequality." *Journal of Political Economy* 111, no. 4: 818–57.

CBO (Congressional Budget Office). 1998. *Social Security and Private Saving: A Review of the Empirical Evidence*. Washington: CBO. https://www.cbo.gov/sites/default/files/105th-congress-1997-1998/reports/ssprisav.pdf.

———. 2019. *The Budget and Economic Outlook: 2019 to 2029–January 2019*. Washington: CBO. https://www.cbo.gov/system/files/2019-03/54918-Outlook-3.pdf.

Christensen, Jens H. E., and Glenn D. Rudebusch. 2019. "A New Normal for Interest Rates? Evidence from Inflation-Indexed Debt." *Review of Economics and Statistics*, forthcoming.

Claessens, Stijn, and M. Ayhan Kose. 2017. "Asset Prices and Macroeconomic Outcomes: A Survey." Working Paper 8259. Washington: World Bank. http://documents.worldbank.org/curated/en/599841511879164538/Asset-prices-and-macroeconomic-outcomes-a-survey.

Constâncio, Vítor. 2016. "The Challenge of Low Real Interest Rates for Monetary Policy." Lecture given at Macroeconomics Symposium at Utrecht School of Economics, Utrecht, June 15. https://www.ecb.europa.eu/press/key/date/2016/html/sp160615.en.html.

D'Amico, Stefania, Don H. Kim, and Min Wei. 2018. "Tips from TIPS: The Informational Content of Treasury Inflation-Protected Security Prices." *Journal of Financial and Quantitative Analysis* 53, no. 1: 395–436.

Deaton, Angus. 1991. "Saving and Liquidity Constraints." *Econometrica* 59, no. 5: 1221–48.

Del Negro, Marco, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti. 2018. "Global Trends in Interest Rates." Staff Report 866. New York: Federal Reserve Bank of New York. https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr866.pdf.

Diamond, Peter A. 1965. "National Debt in a Neoclassical Growth Model." *American Economic Review* 55, no. 5, pt. 1: 1126–50.

Draghi, Mario. 2016. "The International Dimension of Monetary Policy." Speech given at ECB Forum on Central Banking, Sintra, June 28. https://www.ecb.europa.eu/press/key/date/2016/html/sp160628.en.html.

Dynan, Karen E., Jonathan Skinner, and Stephen P. Zeldes. 2004. "Do the Rich Save More?" *Journal of Political Economy* 112, no. 2: 397–444.

ECB (European Central Bank). 2018. "The Natural Rate of Interest: Estimates, Drivers, and Challenges to Monetary Policy." Edited by Claus Brand, Marcin Bielecki, and Adrian Penalver. Occasional Paper 217. Frankfurt: European Central Bank. https://www.ecb.europa.eu/pub/pdf/scpops/ecb.op217.en.pdf.

Eggertsson, Gauti B., Ragnar E. Juelsrud, Lawrence H. Summers, and Ella Getz Wold. 2019. "Negative Nominal Interest Rates and the Bank Lending Channel." NBER Working Paper 25416. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25416.

Eggertsson, Gauti B., Neil R. Mehrotra, and Jacob A. Robbins 2019. "A Model of Secular Stagnation: Theory and Quantitative Evaluation." *American Economic Journal: Macroeconomics* 11, no. 1: 1–48.

Elmendorf, Douglas W., and N. Gregory Mankiw. 1999. "Government Debt." In *Handbook of Macroeconomics*, edited by John B. Taylor and Michael Woodford. Amsterdam: North-Holland.

Engen, Eric M., and Jonathan Gruber. 2001. "Unemployment Insurance and Precautionary Saving." *Journal of Monetary Economics* 47, no. 3: 545–79.

Engen, Eric M., and R. Glenn Hubbard. 2004. "Federal Government Debts and Interest Rates." In *NBER Macroeconomics Annual*, edited by Mark Gertler and Kenneth Rogoff. Cambridge, Mass.: MIT Press.

Epstein, Larry G., and Stanely E. Zin. 1991. "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis." *Journal of Political Economy* 99, no. 2: 263–86.

Evans, Paul. 1987. "Interest Rates and Expected Future Budget Deficits in the United States." *Journal of Political Economy* 95: 34–58.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. 2016. "Heterogeneity and Persistence in Returns to Wealth." NBER Working Paper 22822. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w22822.

Fagereng, Andreas, Martin Holm, Benjamin Moll, and Gisle Natvik. 2019. "Saving Behavior across the Wealth Distribution." Working paper. http://www.princeton.edu/~moll/SBWD.pdf.

Faini, Riccardo. 2006. "Fiscal Policy and Interest Rates in Europe." *Economic Policy* 21, no. 47: 444–89.

Farhi, Emmanuel, and François Gourio. 2018. "Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia." *Brookings Papers on Economic Activity*, Fall, 147–223.

Feldstein, Martin. 1974. "Social Security, Induced Retirement, and Aggregate Capital Accumulation." *Journal of Political Economy* 82, no. 5: 905–26.

Feldstein, Martin, and Anthony Pellechio. 1979. ‘Social Security and Household Wealth Accumulation: New Microeconometric Evidence.” *Review of Economics and Statistics* 61, no. 3: 361–68.

Fischer, Stanley. 2016. “Why Are Interest Rates So Low? Causes and Implications.” Speech given at Economic Club of New York, New York, October 17. https://www.federalreserve.gov/newsevents/speech/fischer20161017a.htm.

Gagnon, Etienne, Benjamin K. Johannsen, and David Lopez-Salido. 2016. “Understanding the New Normal: The Role of Demographics.” Working Paper 2016-080. Washington: Federal Reserve Board of Governors. http://dx.doi.org/10.17016/FEDS.2016.080.

Gale, William G., and Peter R. Orszag. 2003. “The Economic Effects of Long-Term Fiscal Discipline.” Working Paper 8. Washington: Urban-Brookings Tax Policy Center. https://www.urban.org/sites/default/files/publication/58901/310669-The-Economic-Effects-of-Long-Term-Fiscal-Discipline.pdf.

Galí, Jordi. 2008. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*. Princeton, N.J.: Princeton University Press.

Gertler, Mark. 1999. “Government Debt and Social Security in a Life-Cycle Economy.” Carnegie-Rochester Conference Series on Public Policy 50: 61–110.

Gordon, Robert J. 2016. “Perspectives on the Rise and Fall of American Growth.” *American Economic Review* 106, no. 5: 72–76.

Guvenen, Fatih. 2009. “An Empirical Investigation of Labor Income Processes.” *Review of Economic Dynamics* 12, no. 1: 58–79.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2015. “What Do Data on Millions of U.S. Workers Reveal About Life-Cycle Earnings Risk?” NBER Working Paper 20913. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w20913.

Hamilton, James D., Ethan S. Harris, Jan Hatzius, and Kenneth D. West. 2016. “The Equilibrium Real Funds Rate: Past, Present, and Future.” *IMF Economic Review* 64, no. 4: 660–707.

Holston, Kathryn, Thomas Laubach, and John C. Williams. 2017a. “Documentation of *R* Code and Data for Measuring the Natural Rate of Interest: International Trends are Determinants.” https://usermanual.wiki/Pdf/HLWCodeGuide.843689289.pdf.

———. 2017b. “Measuring the Natural Rate of Interest: International Trends and Determinants.” *Journal of International Economics* 108, no. S1: S59–S75.

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

Kaplan, Greg, Giovanni L. Violante,, and Justin Weidner. 2014. “The Wealthy Hand-to-Mouth.” *Brookings Papers on Economic Activity*, Spring, 77–138.

Kay, John. 2015. *Other People’s Money: Masters of the Universe or Servants of the People?* London: Profile Books.

King, M., and D. Low. 2014. “Measuring the ‘World’ Real Interest Rate.” NBER Working Paper 19887. Cambridge, Mass.: National Bureau of Economic Research. http://www.nber.org/papers/w19887.pdf.

Kinoshita, Noriaki. 2006. "Government Debt and Long-Term Interest Rates." IMF Working Paper WP/06/63. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Government-Debt-and-Long-Term-Interest-Rates-18866.

Kocherlakota, Narayana. 2015. "Public Debt and the Long-Run Neutral Real Interest Rate." Speech given at Bank of Korea Conference, Seoul, August 19. https://www.minneapolisfed.org/news-and-events/presidents-speeches/public-debt-and-the-long-run-neutral-real-interest-rate.

Kuroda, Haruhiko. 2016. "The Battle against Deflation: The Evolution of Monetary Policy and Japan's Experience." Speech given at Columbia University, New York, April 13. https://www.boj.or.jp/en/announcements/press/koen_2016/data/ko160414a1.pdf.

Laforte, Jean-Philippe, and John Roberts. 2014. "November 2014 Update of the FRB/US Model." *FEDS Notes*. https://doi.org/10.17016/2380-7172.0036.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112, no. 2: 443–77.

Laubach, Thomas. 2009. "New Evidence on the Interest Rate Effects of Budget Deficits and Debt." *Journal of the European Economic Association* 7, no. 4: 858–85.

Laubach, Thomas, and John C. Williams. 2003. "Measuring the Natural Rate of Interest." *Review of Economics and Statistics* 85, no. 4: 1063–70.

Lillard, Lee A., and Yoram Weiss. 1979. "Components of Variation in Panel Earnings Data: American Scientists 1960–70." *Econometrica* 47, no. 2: 437–54.

MaCurdy, Thomas E. 1982. "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis." *Journal of Econometrics* 18, no. 1: 83–114.

Mankiw, N. Gregory. 2000. "The Savers-Spenders Theory of Fiscal Policy." *American Economic Review* 90, no. 2: 120–25.

OECD (Organization for Economic Cooperation and Development). 2018. *OECD Tax Policy Studies: The Role and Design of Net Wealth Taxes*. OECD Tax Policy Study 26. Paris: OECD Publishing. https://www.oecd.org/ctp/the-role-and-design-of-net-wealth-taxes-in-the-oecd-9789264290303-en.htm.

———. 2019. "OECD Tax Database." http://www.oecd.org/tax/tax-policy/tax-database/.

Oh, Hyunseung, and Ricardo Reis. 2012. "Targeted Transfers and the Fiscal Response to the Great Recession." *Journal of Monetary Economics* 59: S50–S64.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Cambridge, Mass.: Belknap Press.

Piketty, Thomas, and Emmanuel Saez. 2007. "How Progressive Is the U.S. Federal Tax System? A Historical and International Perspective." *Journal of Economic Perspectives* 21, no. 1: 3–24.

Plosser, Charles I. 1987. "Fiscal Policy and the Term Structure." *Journal of Monetary Economics* 20: 343–67.

Powell, Jerome H. 2018. "The Federal Reserve's Framework for Monitoring Financial Stability." Speech given at Economic Club of New York, New York, November 28. https://www.federalreserve.gov/newsevents/speech/powell20181128a.htm.

Rachel, Łukasz, and Thomas D. Smith. 2015. "Secular Drivers of the Global Real Interest Rate." Working Paper 571. London: Bank of England. https://www.bankofengland.co.uk/working-paper/2015/secular-drivers-of-the-global-real-interest-rate.

———. 2017. "Are Low Real Interest Rates Here to Stay?" *International Journal of Central Banking* 13, no. 3: 1–42.

Sajedi, Rana, and Gregory Thwaites. 2016. "Why Are Real Interest Rates So Low? The Role of the Relative Price of Investment Goods." *IMF Economic Review* 64, no. 4: 635–59.

Straub, Ludwig. 2017. "Consumption, Saving, and the Distribution of Permanent Income." Working paper, Harvard University. https://scholar.harvard.edu/straub/publications/consumption-savings-and-distribution-permanent-income.

Walsh, Carl E. 1998. *Monetary Theory and Policy*. Cambridge, Mass.: MIT Press.

Wicksell, Knut. 1898. *Interest and Prices*. New York: Sentry Press.

Winter, Christopher. 2016. "The Impact of Government Debt on the Long-Run Natural Real Interest Rate: A Quantitative Evaluation." *Journal of Applied Economics Letters* 24, no. 20: 1429–34.

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, N.J.: Princeton University Press.

Yaari, Menahem E. 1965. "Uncertain Lifetime, Life Insurance, and the Theory of the Consumer." *Review of Economic Studies* 32, no. 2: 137–50.

Comments and Discussion

COMMENT BY

GAUTI B. EGGERTSSON Let me start with an overview of how this paper fits in the modern literature on the liquidity trap—that is, the literature on the zero lower bound (ZLB). I find it useful to separate this into three generations of models. The first-generation models considered the ZLB as being due to some temporary exogenous forces, such as preference shocks. Papers in this vein include two published in this journal: one by Paul Krugman (1998), and my paper with Michael Woodford (2003). The second-generation models instead study the nature of the underlying shocks. Examples include papers by Eggertsson and Krugman (2012) and by Veronica Guerrieri and Guido Lorenzoni (2018). A common culprit for a ZLB episode in the second-generation models is a disturbance in the financial sector, such as a household debt deleveraging shock or shocks to bank balance sheets. A key lesson that emerged from this literature was that most of the proposed shocks predicted a temporary reduction in the natural rate of interest that then recovers—for example, once households put their debt on a more sustainable level or banks clean up their balance sheets. At that point, the ZLB is no longer a constraint on policy and things return to normal. A common theme in the first- and second-generation models is that the central bank can have a big effect on outcomes by managing expectations about how it will conduct interest rate policy once the natural rate of interest has recovered. This has been usually termed "the forward guidance of central banks." It is a very powerful force in these models to limit output contractions.

The third-generation ZLB models emerged after a speech by one of the authors of this paper, Lawrence Summers, at the International Monetary Fund in 2013. The context of his speech was that in 2013, the world's recovery from the financial crisis of 2008 was still anemic, even if, by most accounts, households and bank balance sheets had been put on a more sustainable basis. Despite this normalization, markets projected the interest rate to remain low for a long time, and most people were disappointed by the slow recovery of output from the crisis. Summers (2013a, 2013b) suggested that perhaps the fall in the natural rate of interest was not just a temporary reaction to the financial crisis but instead a permanent drop—in contrast to the explicit or implicit prediction of the first- and second-generation ZLB models. If correct, this insight had broad and far-ranging implications for both modeling and economic policy. For one thing, forward guidance about interest rate policy once the natural rate normalizes is of limited help if there is no prospect of this normalization on the horizon. Overall, this perspective tends to shift the focus away from monetary policy as the major remedy for recessions and tilts the balance more toward fiscal activism, a prescription that has yet to make a serious mark on how actual policy is conducted today, but one that is well aligned with earlier Keynesian thinking.

For the idea of a permanent reduction in the natural rate of interest, Summers used the term "the secular stagnation hypothesis," an expression that originates in Alvyn Hansen's 1938 presidential speech to the American Economic Association (Hansen 1939). At the time of that speech, the United States was experiencing the second phase of the Great Depression. Hansen suggested that the United States might permanently experience sluggish demand, due to a slowdown in population growth and a lack of investment opportunities. Hansen's dire predictions, of course, turned out to be widely off the mark. The baby boom that followed World War II drastically changed the country's population dynamics, and there was a boom in all sorts of innovations. In the postwar prosperity that followed, too much inflation and high interest rates were the worry of the day, not deflation and low interest rates, which are at center stage in the secular stagnation literature. Summers suggested that perhaps Hansen might just have uttered his dire predictions 75 years too early.

Right after Summers's speech, the literature quickly started to formalize these ideas in the modern modeling context.[1] The current paper by Rachel

1. To my knowledge, the paper by Eggertsson and Mehrotra (2014) was the first one to formalize the idea in the modern dynamic stochastic general equilibrium model, using a small-scale overlapping-generations model. Eggertsson, Mehrotra, and Robbins (2019), in addition, considered a quantitative, large-scale overlapping-generations model. In those models, there can be a permanent decline in the natural rate due to a slowdown in population growth, a rise in life expectancy, a rise in productivity, a rise in inequality, and a fall in the relative price of investment. Caballero and Farhi (2018), however, show that a reduction in safe assets supply can also trigger a persistent fall in the natural rate of interest. Eggertsson and others (2016), conversely, consider the open economy dimension of these ideas.

and Summers synthesizes this literature, and extends it in several directions, putting special emphasis on how government policies may have counteracted the fall in the natural rate of interest—for example, by increases in government debt and in pay-as-you-go Social Security. Both policies tend to raise rather than reduce the natural rate of interest. Here, I highlight a few key results of the paper, do some nitpicking, and conclude by highlighting some unresolved questions that have to do with optimal policy responses to secular stagnation.

KEY RESULTS One major result in Rachel and Summers's paper is an empirical estimate of the natural rate of interest in the industrialized economies, suggesting that it has declined by about 300 basis point since 1980, leaving it today in the neighborhood of zero. The authors do a nice extension of a well-known method by Thomas Laubach and John Williams (2003) by considering the industrialized economies as a single bloc. Then they study how much government debt and Social Security have offset downward pressure on interest rates from private sector imbalances. Their results suggest that if government debt had not increased from about 20 to 70 percent of GDP, then the real interest rate would have declined even further, by about 100 basis points, while the increase in Social Security spending led to an additional offset by about 50–100 basis points. The authors thus leave us with the dire prediction that if it had not been for these government policies, then the natural rate today would be at –2 to –3 percent. In this case, today the ZLB would presumably be a severely binding constraint across all advanced economies. This is a very interesting result that goes against the conventional wisdom. An incredible amount of ink has been spilled about the looming crisis due to "unfunded" Social Security entitlement; and, similarly, there is a great deal of alarm over a supposed fiscal crisis that is around the corner due to the rise in government debt over the past decades. Contrary to this view, Rachel and Summers's paper suggests that had it not been for these developments, the industrialized world would currently be mired in a much deeper deflationary crisis than what we already see today in Japan and the euro zone. One takeaway from the empirical estimate that the authors choose not to highlight may seem to be that if low natural interest rates remain a problem in the future, then all the government has to do is to increase government debt and/or pay-as-you-go Social Security spending. As I again discuss at the end of this comment, the solution to secular stagnation may not be that simple.

Although these results are purely empirical, or are imputed using existing studies, the authors go beyond this empirical work, taking the paper from being only interesting to being excellent. We are also presented with two interesting structural models—one with overlapping generations that is ideal for studying the effect of age dynamics, and the other with agents confronted with uninsurable risk, a model ideal for studying inequality. They use these models to account for the empirical finding just highlighted, and in the process they are able to decompose the key drivers of the decline in the natural interest rate. The main conclusion is that, using 1970 as a benchmark (see Rachel and Summers's table 7), the natural rate of interest has dropped by about 320 basis points since then, with the private sector accounting for a drop of 690 basis points and the public sector offsetting this by 360 basis points. The largest individual component contributing to the fall in the natural rate is a fall in total factor productivity and population and working age dynamics (retirement, life expectancy, length of working lift, and the like). Inequality also plays a nontrivial role, accounting for a drop of about 100 basis points. The role of government policy also is roughly in line with the empirical estimates.

THE BIG PICTURE Overall, I interpret the findings, especially the results from the structural models, in a similar way as the findings in a recent paper I wrote jointly with Neil Mehrotra and Jacob Robbins (2019). Like Rachel and Summers, we also find that a model with an overlapping generation structure—although very different in the details of how the age pyramid is modeled—can account for a substantial drop in the real interest rate, from about 3 percent to modestly negative natural rates today in the context of a U.S. calibration.[2] The way I interpreted our finding was that the fall in the real interest rate observed in the data *can* be accounted for by age dynamics and a productivity slowdown and other slow-moving forces. I use the word "can" advisedly because I think that our model—and the same applies to that of Rachel and Summers—leaves enough free parameters so that the results are quite sensitive to assumptions. One could also tell stories consistent with the fact that the fall observed in natural rates is only temporary and due to the global financial crisis and its aftermath, so that they will ultimately rise to a more "normal" level (even if this sort of story becomes increasingly implausible the longer time elapses from the financial crisis and the longer the market seem to predict a low real interest rate). An important next step in the literature is to identify more clearly what elements of the structural models would lead us to one conclusion relative to the other.

This is not a criticism of Rachel and Summers's paper, however. What I think these results show quite conclusively is that a permanent fall in the

2. We have an 80-generation, medium-scale dynamic stochastic general equilibrium model in the tradition of Auerbach and Kotlikoff (1987).

natural rate *is a very plausible scenario that we ought to take seriously,* even if I think it is a bit too early to assign precise probability weights on the secular stagnation scenario relative to others. But the fact that the secular stagnation scenario is very much a plausible possibility should be considered a major result. A few years ago, very few people took seriously the idea that the ZLB would ever become an issue in the United States. They turned out to be very wrong. Today, similarly, I think far too few people have taken seriously the possibility that the ZLB will be a permanent feature of the landscape of stabilization policy in the coming years. Rachel and Summers's paper suggests that studying this uncomfortable possibility should be a first-order priority for macroeconomics. In this respect, their paper should be a wakeup call.

COMMENTS AND CONCERNS I would be remiss as a commenter if I did not do some nitpicking. One of the more interesting experiments the authors undertake is to consider the effect that the increase in inequality has had on the natural interest rate. Inequality rises due to the increase in the variance of the income process of infinitely lived agents that live in the model. The rise in income uncertainty, in turn, increases the precautionary savings of the agents, so they will increase their demand for savings to insure against future negative income shocks. This increase in savings puts downward pressure on the interest rate.

It is not obvious to me that the rise in income inequality we see in the United States has much connection with a rise in idiosyncratic income shocks that people need to insure against. Loosely speaking, it seems instead that the "rich are getting richer and the poor are getting poorer" and that there actually is a decline in mobility "between rich and poor"—that is, if you are born rich, you die rich. Thus, it is not clear that the rise in income inequality is tied to idiosyncratic income risk, for reasons discussed in footnote 3 below; or, in any event, this would need to be established empirically.[3]

3. To be more concrete: Imagine that that there are two types in the model, low type (poor) and high type (rich), and that this generates income inequality. Imagine, now, that there is no transition between types—that is, the poor always stay poor and the rich always stay rich—but allow for some idiosyncratic risk for year type (both rich and poor get sick, and so on). Now if you increase the income of the high types at the expense of the low type without affecting idiosyncratic uncertainty within a group, it is not obvious to me that this has any effect on the interest rate (to see this, you can shut down the idiosyncratic income risk, in which case there is no effect of increasing inequality on interest rates, as both types perfectly smooth consumption and the interest rate is given by the inverse of the discount factor). In the model of Rachel and Summers's paper—where agents are identical, except that they draw idiosyncratic income shocks—income inequality always must lead to an increase in uninsurable income risk, in contrast to the example discussed above.

My impression is that this sort of criticism applies to most of the literature that uses these types of models, and Rachel and Summers discuss this point at some length in their paper. In any event, to the extent that one is considering an infinitely lived household that is constantly drawing new income shocks, my guess is that the motive for precautionary savings is exaggerated. It will be interesting to see future research that merges the life-cycle model the authors present and the model with idiosyncratic risk, where one might be able to get at these issues inside a structural model.

Having said this, I do not think that the authors are necessarily exaggerating the effect of income inequality on the real interest rate. There is a body of literature that emphasizes that the "rich save more" for reasons independent of precautionary motives (see, for example, Dynan, Skinner, and Zeldes 2004). If this is correct, this might work toward an even stronger effect of inequality on the real interest rate than is documented here.

UNRESOLVED QUESTIONS AND POLICY ISSUES I think that one of the goals of this paper is to make the case that a permanent fall in the natural rate of interest is very much a plausible possibility. I think Rachel and Summers's paper succeeds on that score. A major question this finding then raises is what can be done about it.

When the first-generation ZLB models were developed, there was a tendency among academics to minimize the challenge imposed by these models on monetary authorities, with papers being written with titles such as "Japanese Monetary Policy: A Case of Self-Induced Paralysis?" (Bernanke 2000) and "A Foolproof Way Out of Escaping from a Liquidity Trap" (Svensson 2001). The perception was that the problem posed by the ZLB was easily solved. After all, central banks could just print money. For example, Kenneth Rogoff commented in this journal on Krugman's 1998 paper—which arguably launched the first generation of ZLB models, I think reflecting a relatively broad professional consensus—that "no one should seriously believe that the BOJ would face any significant technical problems in inflating if it puts its mind to the matter, liquidity trap or no. For example, one can feel quite confident that if the BOJ were to issue a 25 percent increase in the current supply and use it to buy back 4 percent of government nominal debt, inflationary expectations would rise." Since then, of course, the Bank of Japan has not increased its money supply by only 25 percent. Relative to 1998, it has almost increased the monetary base 10 times over. And this without inflation budging!

The reaction to the secular stagnation hypothesis by the economics profession has, in my experience, been somewhat similar and could be summarized similarly to how people reacted to the first-generation ZLB models: "It didn't happen, it will not happen, it cannot happen." And as for secular stagnation, if permanently low interest rates were ever truly a problem, some argue, why would the government simply not increase its debt until the interest rate rises? Is that not an obvious free lunch?

A comparative static that my colleagues and I recently presented puts an interesting perspective on this argument (Eggertsson, Mehrotra, and Robbins 2019). In this paper, we show that in a calibrated model of the U.S. economy, the ratio of debt to GDP would need to increase by over 215 percent of GDP for the interest rate to rise from –1.5 percent to 1 percent. If that sort of increase in the debt is needed in the real world, the question becomes: Could other forces, outside the model, start to play a role that could change the conclusion?

This is a question studied in a recent paper by Vaishali Garga (2019). Garga considers the plausible scenario—likely to be relevant in the real world—that even if the economy finds itself in secular stagnation of the type considered in Rachel and Summers's paper, the public will (for good reasons) put some probability on the fact that the secular stagnation hypothesis will turn out to be incorrect and that the economy will instead transition at some future date into a "normal" state where interest rates are positive again. Her point is that if the debt is high enough, then this reversal to normality must trigger a fiscal crisis with associated tax increases and entitlement cuts. The mere expectation of this scenario, in turn, can then undo the positive effect that increasing the government's debt has on the interest rate during secular stagnation. Higher debt during secular stagnation, in other words, can trigger people to save for the possibility of a fiscal crisis state and possible cuts in their Social Security. Theoretically, she shows that for a positive probability of a reversal of this kind, there will always be a tipping point for government debt, above which increasing debt will lead to a further decline in the natural rate of interest rather than an increase. Though this tipping point surely exists in theory, the question is how high it is empirically. Garga then shows evidence suggesting that in the case of Japan, this force might be quite strong—perhaps even strong enough so that the tipping point would be reached. All this is to say that simply increasing government debt may not be the silver bullet to solve the problem of secular stagnation.

This paper by Rachel and Summers, and the others on which it builds, have in my view conclusively shown that secular stagnation is a plausible scenario. It thus seems all the more urgent to begin doing more research on how this problem can be solved.

REFERENCES FOR THE EGGERTSSON COMMENT

Auerbach, Alan J., and Laurence J. Kotlikoff. 1987. *Dynamic Fiscal Policy*. Cambridge: University Press Cambridge.

Bernanke, Ben S. 2000. "Japanese Monetary Policy: A Case of Self-Induced Paralysis?" In *Japan's Financial Crisis and Its Parallels to the U.S. Experience*, edited by Ryoichi Mikitani and Adam S. Posen. Washington: Petersen Institute for International Economics.

Caballero, Ricardo J., and Emmanuel Farhi. 2018. "The Safety Trap." *Review of Economic Studies* 85, no. 1: 223–74.

Dynan, K., J. Skinner, and S. Zeldes. 2004. "Do the Rich Save More?" *Journal of Political Economy* 112, no. 2: 397–444.

Eggertsson, Gauti B., and Paul Krugman. 2012. "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach." *Quarterly Journal of Economics* 127, no. 3: 1469–1513.

Eggertsson, Gauti B., and Neil R. Mehrotra. 2014. "A Model of Secular Stagnation." NBER Working Paper 20574. Cambridge, Mass.: National Bureau of Economic Research.

Eggertsson, Gauti B., Neil R. Mehrotra, and Jacob A. Robbins. 2019. "A Model of Secular Stagnation: Theory and Quantitative Evaluation." *American Economic Journal: Macroeconomics* 11, no. 1: 1–48.

Eggertsson, Gauti B., Neil R. Mehrotra, Sanjay R. Singh, and Lawrence H. Summers. 2016. "A Contagious Malady? Open Economy Dimensions of Secular Stagnation." *IMF Economic Review* 64, no. 4: 581–634.

Eggertsson, Gauti B., and Michael Woodford. 2003. "The Zero Bound on Interest Rates and Optimal Monetary Policy." *Brookings Papers on Economic Activity* 34, no. 1: 139–234.

Garga, Vaishali, 2019. "Fiscal Expansion in a Secular Stagnation: What If It Isn't Secular?" Working paper. https://www.vaishaligarga.com/research.

Guerrieri, Veronica, and Guido Lorenzoni. 2017. "Credit Crises, Precautionary Savings, and the Liquidity Trap." *Quarterly Journal of Economics* 132, no. 3: 1427–67.

Hansen, Alvin. 1939. "Economic Progress and Declining Population Growth." *American Economic Review* 29, no. 1: 1–15.

Krugman, Paul R. 1998. "It's Baaack: Japan's Slump and the Return of the Liquidity Trap." *Brookings Papers on Economic Activity* 29, no. 2: 137–205.

Laubach, Thomas, and John C. Williams. 2003. "Measuring the Natural Rate of Interest." *Review of Economics and Statistics* 85, no. 4: 1063–70.

Summers, Lawrence H. 2013a. "Speech given at IMF Fourteenth Annual Research Conference in Honor of Stanley Fischer," Washington, November 8. http://larrysummers.com/imf-fourteenth-annual-research-conference-in-honor-of-stanley-fischer/.

———. 2013b. "Why Stagnation Might Prove to Be the New Normal." *Financial Times*, December 15. http://larrysummers.com/commentary/financial-times-columns/why-stagnation-might-prove-to-be-the-new-normal/.

Svensson, Lars E. O. 2001. "The Zero Bound in an Open Economy: A Foolproof Way of Escaping from a Liquidity Trap." *Monetary and Economic Studies* 19, no. S1: 277–312.

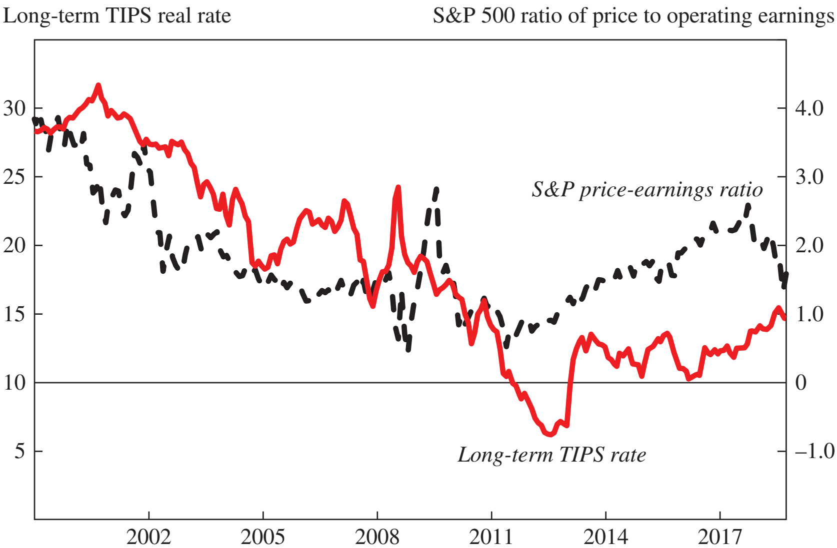
COMMENT BY

ARVIND KRISHNAMURTHY Over the past 30 years, real interest rates in the developed world have fallen by about 3 percent. The core of Rachel and Summers's paper is in their figure 11, which decomposes the drop in the real rate to various economic mechanisms. The remarkable lesson from the paper's analysis, which is evident in the figure, is that there are strong two-way forces at work in determining interest rates. For example, changes in demographics have by themselves led to a fall in interest rates of nearly 3 percent. However, the large rise in government debt has led to an offsetting rise in interest rates of about 1.5 percent. The paper goes through more economic mechanisms, and the broad lesson is that the net of these individually large economic mechanisms has led to a fall in observed real rates of about 3 percent.

The paper comes to these conclusions through the lens of two calibrated models: life-cycle and precautionary savings. The first is a life-cycle model where demographics and savings for retirement determine interest rates. The second model is a Bewley model with idiosyncratic income risk and an insurance channel for determining interest rates. Rachel and Summers's figure 11 is derived from running various experiments within these calibrated models.

In this comment, I argue that a nontrivial portion of the fall in interest rates is due to a decline in the rate of return on safe assets, not all assets. This point is a theme of the recent literature on safe assets. See, in particular, the work of Ricardo Caballero and his coauthors (Caballero, Farhi, and Gourinchas 2008; Caballero and Krishnamurthy 2009). Rachel and Summers acknowledge this point in their paper, but argue that it is a second-order effect. The models they write down are ones where their considered economic channels move the rates of return on all assets equally and not the rates of return particularly on safe assets. I argue the case that the fall in safe rates is not a second-order issue; and to make this point, I offer a set of "maximal" computations tracing the fall in safe rates. I acknowledge at the outset that my computations are subject to considerable uncertainty, but it

Figure 1. S&P 500 Ratio of Price to Operating Earnings and Long-Term TIPS Real Rate, 1999–2018[a]



Sources: Federal Reserve; Bloomberg; author's computations.
a. S&P = Standard & Poor's; TIPS = Treasury Inflation-Protected Securities.

should be clear that for the issue at hand, the authors' computations are also subject to considerable uncertainty.

STOCK MARKET VALUATIONS My figure 1 plots movements in the long-term yield on Treasury Inflation-Protected Securities (TIPS) (the solid line) from 1999 to 2018 as well as the ratio of price to operating earnings for all companies in the Standard & Poor's (S&P) 500 index (the dashed line). The U.S. Treasury first issued TIPS in 1999, which is the start of the sample, and the yield represented corresponds to the longest TIPS bond issued at that time (a 30-year security). In 2003, the Treasury moved to regularly issuing a 10-year TIPS, and that is the yield that is tracked from 2003 onward.

My figure 1 illustrates the sizable fall in real interest rates studied by Rachel and Summers's paper. I ask whether this fall is equally reflected in the discount rate pricing equities. To answer this question, I use the Gordon growth formula:

$$P_t = \sum_{s=t}^{\infty} \frac{\phi E_t (1+g)^{s-t}}{(1+r+ERP)^{s-t}}, \tag{1}$$

where P_t is stock price, E_t is earnings, ϕ is the payout ratio, g is earnings growth, r is the real rate on the safe asset (that is, the TIPS yields), and ERP is the equity risk premium, so that $r + ERP$ is the discount rate applied to risky corporate earnings. Rewriting this expression, we have the price-to-earnings ratio:

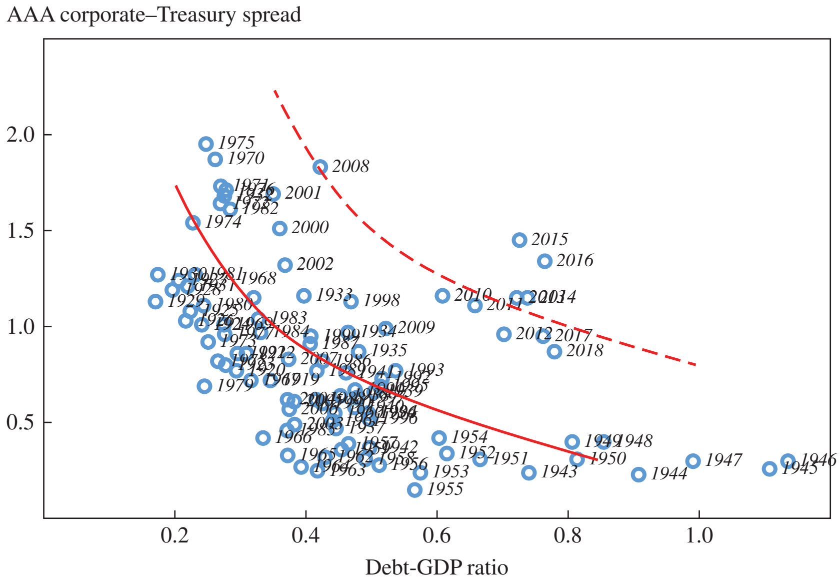
$$\frac{P_t}{E_t} = \frac{\phi}{r + ERP - g}. \qquad (2)$$

We can use this formula to imply the movements in the ERP over the period represented in my figure 1. I compute that the average price/earnings ratio from 1999 to the 2007 precrisis sample is 21.0, and the average interest rate on TIPS is 3 percent. I use a value of g of 3.5 percent, corresponding to precrisis GDP growth, and matching estimates from Rachel and Summers's paper. I also use a payout ratio of 20 percent, but this number does not matter for the main conclusions. Using these calibrations, I solve for the implied ERP, which equals 4.3 percent (to match the 21.0 price/earnings ratio).

Similarly, I compute the ERP for the 2009–18 postcrisis sample. In this sample, the price/earnings ratio averages 17.6, and the TIPS rate averages 0.4 percent. I use a value of 2 percent for g, corresponding to postcrisis GDP growth rates. I keep the payout ratio the same, at 20 percent. The ERP required to match the price/earnings valuation is 6.1 percent. In other words, in the postcrisis sample, the ERP must have risen to rationalize observed price/earnings ratios. Consider a counterfactual. Hold the ERP constant at 4.3 percent, and decrease r and g. The Gordon formula implies that the price/earnings ratio then needs to be 29.6! That is, the fall in interest rates is quite large, and by itself should have led to a sizable increase in the price/earnings ratio. However, the price/earnings ratio over this entire period (ignoring the crisis) is relatively stable. I conclude that the ERP must have risen significantly.

At first glance, this computation flies in the face of financial market commentary. Investors bemoan that stocks are expensive. My computation implies that the expected return on stocks has indeed fallen. That is, the safe rate fell by 2.6 percent (= 3 – 0.4), and the ERP increased by 1.7 percent (= 6.1 – 4.3), indicating that the expected return on stocks has fallen by 0.9 percent. From an investor's standpoint, returns have fallen significantly. In this sense, stocks are expensive. However, on a relative basis, safe bonds have become even more expensive than stocks. The discount rate on all

Figure 2. Corporate Bond Spreads, 1919–2018[a]



Sources: Krishnamurthy and Vissing-Jorgensen (2012), Federal Reserve, author's calculations.

a. This figure shows the 20-year AAA corporate–Treasury spread versus the publicly held Treasury debt-GDP ratio for 1919 to 2018.

assets has fallen by 0.9 percent, while the discount rate on safe assets has fallen an additional 1.7 percent.

My perspective helps to rationalize corporate financing behavior over the last 20 years. As safe rates have fallen, corporate investment has not risen. Flat corporate investment is inconsistent with a decrease in real rates to all assets. But in a world where the risky cost of capital has not fallen as much as the safe rate, we would not expect to see a substantial increase in corporate investment. Indeed, the form of investment that has risen the most over the last 20 years is residential investment. Due to securitization and the banking system, the discount rate to residential investment is a near safe rate. This perspective also helps rationalize the rise in corporate leverage and share buybacks over the last decade. Firms have engaged in a form of capital structure arbitrage: issue safe bonds at low rates, and buy risky stocks, offering higher rates of return.

CORPORATE BOND SPREADS My figure 2 replicates and updates a figure from my (2012) paper with Annette Vissing-Jorgensen. We observe that there is a strong negative relation between the spread of long-term, AAA-rated corporate bonds over long-term Treasury bonds and the outstanding amount of publicly held government debt, normalized by GDP. We interpret the relation (and provide other evidence in support of the interpretation) as reflecting a demand curve, akin to a money demand curve, for the higher safety and liquidity of Treasury bonds relative to corporate bonds. We also present evidence that the safety demand, though reflected most prominently in the movements of Treasury bond yields, extends to high-grade private assets, including illiquid corporate bond rates and money market rates such as bank deposit rates (which underly the determination of interest rate swap rates).

In my figure 2, I have drawn a solid line through the points from 1919 to 2007, roughly indicating the precrisis demand for safe assets. The points in the figure's upper-right quadrant correspond to the points after 2007. It appears from this graph that the demand curve for safe assets has shifted outward over the last decade. Indeed, the vertical distance between the two curves drawn on the figure is about 80 basis points, suggesting an increased safety premium of just under 1 percent (for a given quantity supplied).

There are further observations that reinforce this point. First, consider the left panel of figure 3 in Rachel and Summers's paper. Though all the yields pictured in the figure have fallen over the last 25 years, it is also evident that the spread between the various yields considered has widened over this same period, with Treasury yields falling more than other yields. The widening in the corporate-Treasury spread over the postcrisis period is particularly striking because stock market volatility has also declined significantly over the last decade. The decline is evident in both measured stock return data or in a volatility index such as the Chicago Board Options Exchange's Volatility Index (VIX). From 1995 to 2006, the VIX averaged about 21 percent; and from 2013 to 2018, it averaged about 15 percent. Corporate bonds, as in the classic Robert Merton (1974) model, are a safe bond minus the value of a put option on the underlying firm's assets. As volatility has decreased, we would expect that the value of this put should have fallen. The decreased put value implies a narrowing of the corporate bond to Treasury spread. Empirically, the opposite is true. Corporate bond rates have not fallen toward Treasury rates over this period (see figure 3 in Rachel and Summers's paper). This fact also implies an increased preference for the safety of Treasury bonds. Furthermore, the corporate bond in my figure 2 is an AAA bond, which is very safe. An increase in demand for safe assets would also decrease AAA rates, albeit less so than in Treasury rates. This suggests that the estimates from my figure 2 are an underestimate of the increase in safe asset premia. If one could construct a hypothetical spread between equities and safe Treasury rates, and make a plot as in figure 2, presumably that plot would reflect the larger shift in the demand for all safe assets. In a sense, the price/earnings computation I offered earlier is two points on such a plot. Thus, these two pieces of evidence are internally consistent and reflect a sizable decrease in the return on safe assets.

My figure 2 gives another way to estimate the impact of the expansion of government debt on interest rates. Rachel and Summers estimate from their model that the expansion of government debt from 1970 to the present contributed to an increase in rates of 1.5–2.0 percent. We can empirically check this estimate from my figure 2. Comparing the points in 1970 with the present points, along the same demand curve, indicates a rise of roughly 1.5 percent in rates. From a very different perspective, I arrive at a similar estimate as Rachel and Summers of the impact of the increase in government debt supply on interest rates.

DISCUSSION I conclude that the discount rate on all assets has fallen, consistent with Rachel and Summers's analysis. My analysis also indicates that the discount rate on safe assets has fallen further than the rate on all assets. I have offered a set of maximal computations for the further safe rate decline, estimating this at between 1 and 2 percent.

I do not interpret these computations as invalidating the analysis of Rachel and Summers. Their model is readily interpretable in terms of safe asset demand. A desire for insurance against income risk is likely best met by holding safe assets. Their Bewley model could be easily repurposed to address safe asset rates. Likewise, a demand for retirement savings that is accommodated via defined-benefit pension plans is likely best met by holding safe assets. The life-cycle model could also be repurposed to address safe asset rates.

Thinking about safe asset demand brings other forces to the fore: savings glut and foreign reserve accumulation, collateral and financial intermediation issues such as a shortage of high-quality liquid assets, and risk preferences of investors.

My analysis also does not invalidate the policy conclusions of Rachel and Summers. Low-equilibrium interest rates mean that the zero lower bound will frequently constrain monetary policy. This is because the central bank sets the rate on a safe asset (that is, reserves). Likewise, U.S. government debt is currently the par excellence of safe assets in the world. If low safe asset rates are driven by a high convenience yield on safe assets, then it follows that the government has more fiscal space when these convenience yields are high. Indeed, this fiscal space conclusion is strengthened in the safe asset convenience yield perspective. A Friedman rule–style argument calls for more issuance of convenience assets (government debt) when the convenience yield is high (Friedman 1969). However, there is another counterbalancing force: if the government issues too much debt, such debt may no longer be viewed by investors as safe, and the convenience yield may disappear. Finally, the safe asset perspective identifies further considerations that are relevant for policy. The private sector, in addition to the government, can create safe assets. Financial intermediaries in particular are safe asset creators, as argued by Caballero and Krishnamurthy (2009); Krishnamurthy and Vissing-Jorgensen (2015); and Tri Vi Dang and others (2017). But private safe assets are inherently fragile and can lead to runs and an increase in systemic risk. Thus, if interest rates are low because of high convenience yields on safe assets, policymakers should also be mindful of systemic risk from the private sector.

REFERENCES FOR THE KRISHNAMURTHY COMMENT

Caballero, Ricardo J., Emmanuel Farhi, and Pierre-Olivier Gourinchas. 2008. "An Equilibrium Model of 'Global Imbalances' and Low Interest Rates." *American Economic Review* 98, no. 1: 358–93.

Caballero, Ricardo J., and Arvind Krishnamurthy. 2009. "Global Imbalances and Financial Fragility." *American Economic Review* 99, no. 2: 584–88.

Dang, Tri Vi, Gary Gorton, Bengt Holmström, and Guillermo Ordoñez. 2017. "Banks as Secret Keepers." *American Economic Review* 107, no. 4: 1005–29.

Friedman, Milton. 1969. *The Optimal Quantity of Money, and Other Essays.* Chicago: Aldine.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen. 2012. "The Aggregate Demand for Treasury Debt." *Journal of Political Economy* 120, no. 2: 233–67.

———. 2015. "The Impact of Treasury Supply on Financial Sector Lending and Stability." *Journal of Financial Economics* 118, no. 3: 571–600.

Merton, Robert C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29, no. 2: 449–70.

GENERAL DISCUSSION Valerie Ramey began noting the importance of the paper and asking if the authors had considered whether the rise of China and its relatively recent integration into global markets had an impact on the long-term real interest rate. She referred to a paper by Fernando Broner and others, which documents that the share of government debt held by foreigners has risen dramatically in industrialized countries.[1] For example, the share of U.S. government debt held by foreigners was about 5 percent in 1970 but now sits just below 50 percent.

Martin Baily commended the paper. He stated that the authors' reasons for studying the advanced economies as a bloc is valid, but that doing so may obscure important variation across countries—particularly, the unique situation of the United States. Baily is not convinced that the United States is in a secular stagnation situation like Japan and Europe, whereas Lawrence Summers has argued so in various news outlets and academic papers. He observed that the United States appears to be in excess demand by a certain definition, given that domestic demand appears to satisfy the economy at full employment while net imports are persistently negative. Furthermore, he expressed skepticism that such excess demand is entirely driven by deficit-financed fiscal stimulus at present.

Baily also noted two observations in support of the idea that investment opportunities have declined, in contrast to some of the discussants' claims. The first observation is that the required rate of return on corporate investing is much higher than the risk-free interest rate and does not appear to adjust downward properly when the risk-free interest rate falls. This would imply a lack of real corporate investment opportunities. The second is that many of the world's economies—including those in Latin America and Africa, but also perhaps Japan and those in Europe—have regulatory regimes that create large barriers for would-be investment opportunities.

Jason Furman also commended the paper. He expressed some concern that research on this topic has focused on testing various "hunches" about the causes of the interest rate decline rather than providing meaningful cost-benefit analyses of various policy options in the new interest rate environment. He suggested that a valuable paper would ask the question raised by Gauti Eggertsson in his comment: How much additional government debt is necessary to achieve sufficiently higher interest rates and improved cyclical performance? He suggested that the models developed by John Williams to study the effects of the zero lower bound (ZLB) on interest rates would enable one to perform this kind of cost-benefit analysis. Even if

1. Fernando Broner, Daragh Clancy, Aitor Erce, and Alberto Martin, "Fiscal Multipliers and Foreign Holdings of Public Debt," ESM Working Paper 30 (Luxembourg: European Stability Mechanism, 2018), https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3153111_code2401764.pdf?abstractid=3153111&mirid=1.

it is the case that substantial additional debt only raises the neutral interest rate by a small amount, it remains of interest to policymakers to quantify the trade-offs between countering the effects of the ZLB with various monetary policies or increasing debt to offset low interest rates.

Susanto Basu discussed the interpretation of the decline in total factor productivity (TFP) growth, which was raised by both the authors and commenters. Although figure 4 in the paper and observations by commenters suggest that the *decline* in TFP growth has pushed the neutral interest rate down by about 2 percentage points, Basu interpreted the authors' discussion to suggest that *increases* in TFP growth would result in lower real rates. The authors' model incorporates channels that could cause real interest rates to rise or fall in response to higher TFP growth via agents' expectations of future consumption growth and agents' saving out of current income, respectively. He asked the authors to comment on why TFP's positive relation with real interest rates appears to dominate in the model.

Donald Kohn remarked that the paper helps enlighten why the neutral rate—R^*—appeared to drop sharply during the financial crisis and has not recovered. He noted that the persistence of excess savings since the crisis can help explain this observation, but that a similar balance of (excess) savings and low investment appeared in the 1980s concurrent with a much higher value of R^*. This suggests that the balance of savings is not the only explanatory factor and that perhaps demand for capital has declined significantly since that time.

Kohn noted that the relation between productivity growth and R^* becomes clearer after distinguishing between the long and short runs. In the long run, productivity gains should be expected to raise R^*; in the short run, productivity gains shift the Phillips curve outward, enabling the central bank to promote higher employment via lower interest rates without fueling inflation.

Robert Gordon noted that the paper attributes the steep decline in R^* to declining population and productivity growth, but that consensus estimates of potential output growth from—say, those of the Congressional Budget Office and the Federal Reserve—do not reflect such steep declines in these factors. One could make the case that potential output has fallen substantially more than that of the consensus, however. For example, comparing realized output growth since 2009 with the implied output growth derived from a simple Okun's law (with a coefficient of one-half on the unemployment gap) would yield an estimate of potential output of about 1 percent, not the 1.9 percent that is currently estimated by the Congressional Budget Office. In addition, Gordon observed that potential growth in labor force participation declined from about 1½ percent each year between the 1960s and 1990s, when females were rapidly joining the labor force, to about ½ percent each year in the last decade. Such a decline would imply that potential GDP growth has fallen from about 3 percent to about 1 percent. This decline also appears to have been concentrated between 2007 and 2012 (as deduced from a formal procedure using a Kalman filter that removes cyclical variation in the unemployment gap). The timing thus tracks closely with the timing of a sharp decline in R^* over the same period.

Gordon also emphasized the role of productivity growth in determining R^*. He observed that in any kind of Solow growth model with steady state ratios of capital to output, a decline in potential output growth produces a decline in steady state investment, implying a strong link between potential output growth and the natural interest rate. He emphasized further that causality is more likely to run from low-productivity growth to lower investment, rather than vice versa.

William Brainard noted that the authors' specification of the investment–saving curve assumes the response of output to the interest rate is the same throughout the sample period, whereas R^* time varies, following a random walk. Given the assumption that output responds to the difference between the current interest rate and R^*, any cyclical or trend variation in the response of output to the interest rate (a_r) will affect the contemporaneous Kalman filter estimate of R^*. There are good reasons to expect time variation in a_r, just as there are for R^*. Perhaps most relevant for this paper, housing investment, one of the most interest-sensitive components of demand during tightening, was likely to be insensitive to rate reductions during the Great Recession, when house prices were well below replacement cost. It would be hard to identify both time-varying shocks to a_r and R^*, but if $a_r > 0$ is smaller during the Great Recession and larger during booms, he conjectured that estimated R^* would be lower during the Great Recession and higher during booms than the authors' estimates.

Brainard also remarked that U.S. interest rates are affected by two factors not included in the theoretical models: the important roles that U.S. Treasuries have acquired as a source of collateral in international financial markets and as a safe haven. During the global financial crisis, which began in the United States, global investors fled to U.S. Treasuries for safety, a factor that helps explain low U.S. rates.

Laurence Meyer observed that there are two common views of government deficits at present. The first is that persistent deficits are unsustainable and generate serious risks. The second view, represented by the conclusions of Rachel and Summers's paper, states that higher deficits raise R^* and offset secular declines in other real variables, on net benefiting the economy. He noted that such conclusions appear to be similar to those of modern monetary theory, whose models are not well regarded by the authors or many in the profession. However, the conclusions do suggest that fiscal policy in advanced economies has performed spectacularly over the last few decades, in particular by reducing risks associated with the ZLB. Meyer said the paper was fascinating and provided a very important decomposition of the forces affecting R^*.

Giorgio Primiceri commented that a forthcoming paper by Kurt Lunsford and Kenneth West studies the relationship between TFP growth and the secular trend in the natural rate from 1880 onward. They find that the positive correlation emerges only after 1970.[2]

Robert Hall said he was surprised by Rachel and Summers's confidence in finding a dramatic departure from Ricardian neutrality—the notion that government borrowing decisions are neutral in equilibrium. Nonneutrality, which produces the relationship between government borrowing and the real interest rate, stems from the introduction of finitely lived agents into their paper's central model. However, he observed that saving has become significantly more concentrated among the wealthy in the last few decades, and that the wealthy are much less likely to increase consumption in anticipation of death. That wealth is concentrated among those who pass on their wealth to future generations makes the introduction of Ricardian nonneutrality a puzzling finding.

James Stock added that the decomposition of forces acting on the real rate is very nice, but ought to include the liquidity and safety premiums that Arvind Krishnamurthy described. He observed that there remains about 1.3 percentage points of unexplained decline in R^* over the observed period, which liquidity and safety demand may reasonably explain without contradicting the main findings of the paper.

Lawrence Summers thanked the participants. He stated that the paper's main finding, that the neutral rate has been driven downward by strong secular forces more than its observed decline reveals, remains valid even after incorporating every criticism provided in the current discussion. In addition, the common position held by Arvind Krishnamurthy and others that the decline mostly reflects the attractive liquidity and safety properties of government debt also supports the paper's main policy implications. In

2. Kurt G. Lunsford and Kenneth D. West, "Some Evidence on Secular Drivers of U.S. Safe Real Interest Rates," *American Economic Journal: Macroeconomics*, forthcoming.

other words, if demand for government bonds is so high, the government ought to satisfy this demand and increase deficits.

Summers said that Gauti Eggertsson's comments on incorporating income uncertainty and the potential effects of rising income inequality are correct and should be incorporated into the analysis. Summers acknowledged three important missing components of the intertemporal framework in the paper: the high propensity to save among the rich, the tendency for monopoly power to reduce investment, and, most important, reduced investment opportunities in recent decades. Summers called this last force a "demassification" of the economy, which is the result of several secular and preference-driven trends. He described, for example, that a cell phone has more computing power than a supercomputer used to; that people prefer small apartments in cities to big houses in suburbs; that law firms need less office space per lawyer; and that nobody wants to build a shopping mall. He remarked that he is not strongly convinced that safety and liquidity preferences will explain the residual in the decomposition of the neutral rate.

With regard to the discussion about equity risk premia, Summers noted that the question demands more attention in the paper, but that he does not perceive it to be as important as Krishnamurthy does, for four reasons. First, the fact that the risk premium appeared to be low in the late 1990s and early 2000s likely reflects price bubbles in those periods and not underlying risk preferences. Second, there is no clear reason why a decline of 2 percentage points in the risk premium over that period, if it did occur, should explain a decline of 2 percentage points in the safe rate. Summers said he would benchmark any estimated change in the risk premium to relate to half that change in the safe rate. Third, the observation that yields on AAA-rated bonds increased after the global financial crisis does not reflect structural changes in safety preferences as much as it does a reassessment of the asset class following high default rates during the crisis. Fourth, he observed that the opinion that the safe rate has declined because the risk premium has risen comes with an important corollary: demand for stocks should be significantly higher than for bonds right now. This view remains far from the prevailing opinion of many who study this topic.

Summers further replied to comments about the importance of safety and liquidity premia that the co-movement across a variety of asset classes—including assets that were once regarded to be severely illiquid, like inflation-protected Treasuries—suggests that the secular forces of focus in the paper are important. Though there may be some role for safety demand, he said, he remains confident in the choice to focus on an intertemporal saving and consumption framework.

He replied to Valerie Ramey's comment by observing that the global current accounts surplus of the industrial countries is roughly zero and has even risen slightly in recent years, implying that any foreign savings demand would be captured in their analysis by treating the advanced economies as a bloc.

Summers agreed with Martin Baily's comment that policies in the rest of the world that increase investment opportunities would raise the advanced economy normal rate. On Baily's claim that the United States is not likely experiencing secular stagnation, he replied that the comment is likely true under the assumption that the U.S. is a closed economy. He agreed that there appears to be some excess demand in the U.S., but noted that estimating the neutral rate under the assumption the U.S. is an open economy is not straightforward but relies on the level of the exchange rate. For example, if the U.S. were to choose to close its current account deficit, the dollar would depreciate against other advanced currencies, and Europe and Japan would face even more severe stagnation and lower interest rates. He emphasized that the analysis examines the industrial world, not only the United States, precisely for this reason.

Summers agreed with Jason Furman that examining the relative costs of having excessive government debt versus excessively low interest rates is important. And Summers acknowledged Susanto Basu's observation that the discussion of the relationship between TFP growth and interest rates in the paper contradicted that in his own statements. He noted that his paper with Christopher Carroll found, across countries, that as people expect to have rising incomes, they do not reduce their savings rates, in line with model predictions.[3] It may be that people rely on saving habits and are slow to adjust their consumption in response to expected future income shocks.

He acknowledged that comments from Donald Kohn and Robert Hall about the timing of sharp declines in R^* are interesting, but that he does not take the sharp declines too seriously, for two reasons. The first is because of the kind of structural investment changes that occurred in the global financial crisis and its aftermath, as William Brainard observed; the second is that the estimates that reveal that sharp downturn in the United States are

3. Christopher D. Carroll and Lawrence H. Summers, "Consumption Growth Parallels Income Growth: Some New Evidence," in *National Saving and Economic Performance*, edited by B. Douglas Bernheim and John B. Shoven (Chicago: University of Chicago Press, 1991), http://econ.jhu.edu/people/ccarroll/papers/CParallelsY.pdf.

found using closed-economy models. These estimates are dubious, he noted, because the U.S. has a fluctuating current account.

Summers thought Brainard's proposal that the slope of the investment-saving curve had shifted after the financial crisis was interesting, and that if he were right, it would help explain why various fiscal effects are now estimated to have slightly larger effects on neutral interest rates.

Summers also replied to Laurence Meyer, who had noted that the paper's conclusion may be construed as support for inflation-financed debt expansion, that he has made clear that he favors a balanced approach to fiscal expansion. He referenced his *Foreign Affairs* article with Jason Furman that outlines the reasons for this balanced approach and potential policy options.[4] He emphasized that he finds the idea that printing money reduces the cost of running deficits, which is the central idea in modern monetary theory, to be nonsense.

Summers noted that he does not find Lunsford and West's evidence on the historical link between TFP growth and neutral interest rates to be especially surprising or important.

He replied to Robert Hall that Ricardian neutrality may break down for many reasons beyond the consumption curvature across the life cycle—for example, simply because consumers are not sophisticated enough to properly reorganize consumption in response to changes in expected national debt. Furthermore, under many versions of the Ricardian Equivalence assumption, the basic conclusion that fiscal policy can afford to be expansionary at present would still hold, as the creation of government liabilities would create safe assets and interest rates would remain unchanged.

4. Jason Furman and Lawrence H. Summers, "Who's Afraid of Budget Deficits?" *Foreign Affairs*, March–April 2019, https://www.foreignaffairs.com/articles/2019-01-27/whos-afraid-budget-deficits.

WEI CHEN
Chinese University of Hong Kong

CHANG-TAI HSIEH
University of Chicago

XILU CHEN
Chinese University of Hong Kong

ZHENG SONG
Chinese University of Hong Kong

A Forensic Examination of China's National Accounts

ABSTRACT China's national accounts are based on data collected by local governments. However, because local governments are rewarded for meeting growth and investment targets, they have an incentive to skew local statistics. China's National Bureau of Statistics (NBS) adjusts the data provided by local governments to calculate GDP at the national level. The adjustments made by the NBS have averaged about 5 percent of GDP since the mid-2000s. On the production side, the discrepancy between local and aggregate GDP is entirely driven by the gap between local and national estimates of industrial output. On the expenditure side, the gap is in investment. Local statistics increasingly misrepresent the true numbers after 2008, but there was no corresponding change in the adjustment made by the NBS. Using publicly available data, we provide revised estimates of local and national GDP by reestimating the output of firms in the industrial, construction, and wholesale and retail trade sectors, using data on value-added taxes. We also use several local economic indicators that are less likely to be manipulated by local governments to estimate local and aggregate GDP. These estimates also suggest that the adjustments by the NBS have been insufficient since 2008. Relative to the official numbers, we estimate that GDP growth from 2010 to 2016 was about 1.8 percentage

Conflict of Interest Disclosure: Chang-Tai Hsieh is the Phyllis and Irwin Winkelried Professor of Economics at the University of Chicago. Zheng (Michael) Song is a professor in the Department of Economics of the Chinese University of Hong Kong, where Wei Chen and Xilu Chen are also PhD students. Beyond these affiliations, Song received financial support from the General Research Council of Hong Kong. None of the authors received 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 the views of the University of Chicago or the Chinese University of Hong Kong.

points lower and that the investment and savings rate in 2016 was about 7 percentage points lower.

China's national accounts are primarily based on data collected by local officials. However, as documented by Wei Xiong (2018), local officials are rewarded for meeting growth and investment targets. Therefore, it is not surprising that local governments have an incentive to skew the statistics on local growth and investment. The statistical agency of the Chinese government, the National Bureau of Statistics, attempts to correct this bias using administrative data and other sources of data that it gathers directly. The accuracy of the final numbers of aggregate GDP and its components depends on the extent of misreporting by local officials, the data that NBS has at its disposal to correct the misreporting, and the effort it undertakes to do so.

Local GDP is measured via the production approach from three major surveys—of large industrial sector firms, large service sector firms, and "qualified" construction firms. These data are supplemented with surveys of smaller industrial firms and administrative data from other government departments to obtain a number for local GDP on the production side. On the expenditure side, local officials provide estimates of local consumption, investment, government spending, and net exports (vis-à-vis other localities in China and other countries). The two main sources are surveys of household income and expenditures (similar to the U.S. Consumer Expenditure Survey), from which they estimate local consumption, and survey data on investment projects, from which they estimate local investment. Because the sum of local consumption and investment typically exceeds local GDP measured on the production side, the remainder is attributed to local net exports.

The NBS does not simply add up the statistics reported by local governments to arrive at the national aggregates. The NBS also has access to the micro data of the surveys used by local governments, and it supplements these data with economic censuses and administrative data on such categories as land sales, vehicle registration, financial transactions, and foreign trade. Based on these data, the NBS produces its own numbers for national GDP and its components on the production and expenditure sides. The adjustment made by the NBS to the local statistics can be seen by the discrepancies between local GDP and national GDP.

In this paper, we check which of the numbers provided by local governments differ from their national counterparts and, hence, are likely to be inaccurate. First, we show that the sum of local GDP frequently exceeds national GDP. Second, we compare the sum of the local consumption, investment, and net exports with national consumption, investment, and net exports reported by the NBS. We find little discrepancy between local and national consumption but large discrepancies between local and national statistics on investment and net exports. Third, we compare the sum of value-added taxes of sectors as reported at the local level with the same sectors at the national level. We find large discrepancies for the industrial sector and smaller gaps for the nonindustrial sectors.

We then use two approaches to determine the accuracy of adjustments to the local numbers made by the NBS. First, we adjust national GDP by the difference between value-added tax growth reported by the NBS and value-added tax revenue growth reported by the State Administration of Taxation in the sectors where the value-added tax is a major type of taxation. Our estimate suggests that the adjustments made by the NBS were roughly accurate until 2007–8 but that the adjustments made after this date no longer appear to be accurate. Our baseline estimate of GDP growth from 2010 to 2016 is about 1.8 percentage points lower than the official growth rate. Furthermore, our estimate of the aggregate investment and savings rate in 2016 is about 7 percentage points lower than the official numbers.

We use this same approach to adjust local production and expenditure GDP for each Chinese province. There is a positive relation between our adjustments to local GDP and investment across provinces. This evidence suggests that local governments inflate local GDP by overestimating both local production and local investment.

A second approach is to estimate a statistical model where we estimate the relationship between a set of economic indicators (which are less likely to be manipulated) and local GDP before 2008. We then use parameters of the estimated model, along with the same set of the indicators after 2008, to predict local GDP after 2008. The indicators include satellite night lights, national tax revenue, electricity consumption, railway cargo flow, exports, and imports. We use the method developed by Liangjun Su, Zhentao Shi, and Peter Phillips (2016) to control for hidden economic structural heterogeneities across regions. Using this method, we also find that the corrections made to national GDP no longer appear appropriate after 2008. Encouragingly, the adjustments to local GDP made using the two approaches are highly correlated. This provides additional support for our adjustments.

Our revised numbers for the Chinese national accounts thus indicate that the slowdown in Chinese growth since 2008 has been more severe than suggested by the official statistics. At the same time, the true savings rate probably declined by about 11 percentage points from 2009 to 2016, with more than 80 percent of the savings decline showing up in the investment rate and the remainder in the external surplus. In this sense, our revised numbers for China's national accounts also indicate that Chinese growth is associated with consumption growth rather than investment and external surpluses.

I. China's GDP Accounting System

The Chinese national and local GDP statistics are compiled separately. Local statistical bureaus provide estimates of local GDP and its components on the production and expenditure sides. The NBS uses the same data collected and used by local governments, along with the data it collects independently, to arrive at a number for national GDP. The number provided by the NBS is the "official" number for Chinese GDP.

Although the local statistical bureaus are de jure branches of the NBS and are supposed to follow the statistical procedures set by the NBS, de facto they are branches of local governments. The budgets of the local statistical bureaus come from local governments, and officials of local statistical bureaus are evaluated and promoted by local governments. Because of this structure, local statistical bureaus are susceptible to pressure from local officials, who may have an incentive to report inaccurate statistics. The NBS is aware of this bias, and it therefore adjusts the numbers of local GDP provided by the local statistical bureaus.

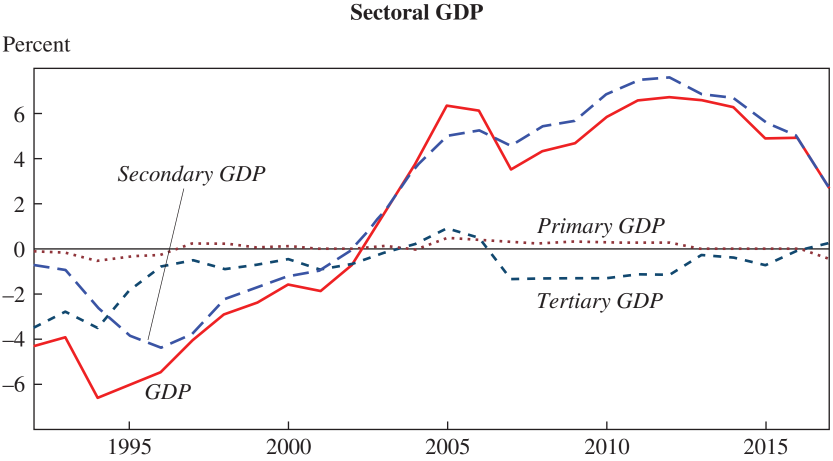
To assess the quality of the official numbers for local and national GDP, we proceed in three steps. First, we compare the sum of local GDP with aggregate GDP provided by the NBS (hereafter, we use the term "aggregate GDP" to refer to the number provided by the NBS). Second, we assess the data used to estimate GDP on the production side. And third, we assess the data used to construct GDP on the expenditure side.

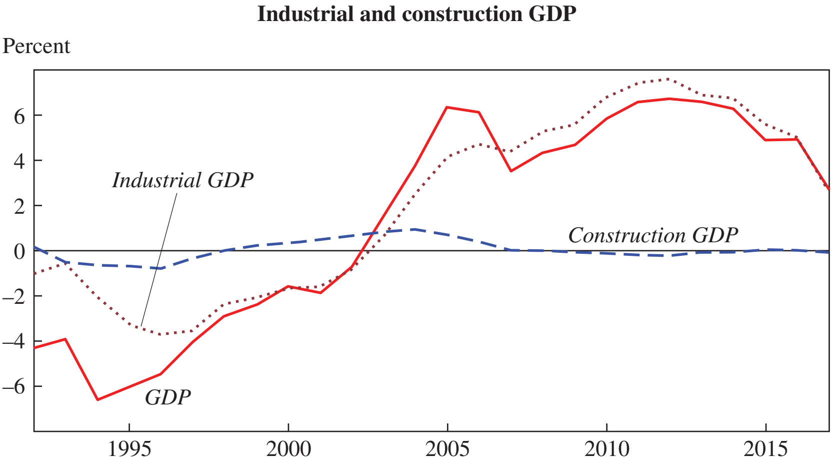
I.A. Comparing Local GDP with Aggregate GDP

The solid line in the top panel of figure 1 shows the magnitude of the adjustment made by the NBS to the local statistics.[1] The figure shows

1. All the national accounts data between 1993 and 2017 were extracted from the NBS website on December 10, 2018. Some numbers—GDP in the primary and tertiary sectors in 2007–16—were updated in February 2019. The changes are very small, and our results are essentially unchanged with the updated numbers. The 1992 provincial data are from Hsueh and Li (1999).

Figure 1. The Gap between Local and Aggregate GDP, by Sector, 1992–2017[a]





Source: National Bureau of Statistics of China.

a. This figure plots the difference between the sum of local sectoral GDP and aggregate sectoral GDP as a percentage of the aggregate GDP for each sector.

the gap between the sum of GDP of each province and aggregate GDP provided by the NBS as a percentage of aggregate GDP. Local governments *understated* GDP relative to the NBS in the 1990s. The sum of local GDP was about 4 to 6 percentage points *lower* than aggregate GDP in the mid-1990s. This pattern changed after 2003. After this date, the sum of local GDP surpassed aggregate GDP and the gap was about

6 percentage points higher than aggregate GDP in 2006. The gap between these two numbers for China's GDP stabilized at about 5 percent of GDP after 2006.

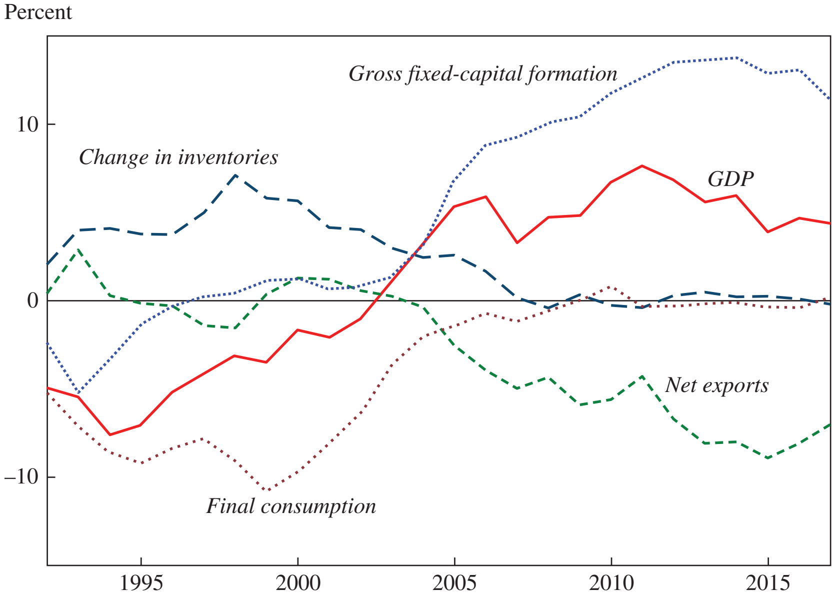
Local statistical authorities and the NBS also provide estimates of local and aggregate GDP by broad sectors. Figure 1 shows the gap between the sum of local GDP and aggregate GDP for each sector as a share of aggregate GDP (for all sectors). The top panel of figure 1 shows the ratio of the sum of local GDP to aggregate GDP for agriculture ("primary"), industry and construction ("secondary"), and services ("tertiary"). Before 2003, the sum of secondary GDP at the local level was lower than aggregate secondary GDP. Furthermore, from about 1997 to 2003 almost all the gap between local and aggregate GDP came from the gap in the industrial sector. Before 1997, some of the gap was due to the discrepancy between local and aggregate statistics for the service sector.

After 2003, all the discrepancy between local and aggregate GDP comes from the industrial sector. The bottom panel of figure 1 shows the comparison of industrial (mining, manufacturing, and public utilities) and construction GDP reported by local governments with that provided by the NBS. As can be seen, the gap between local and national statistics after 2003 is entirely in industry. This finding echoes the research of Carsten Holz (2014) and Ben Ma and others (2014), who also find that the inconsistency between provincial and national GDP mainly came from the industrial sector.

Figure 2 compares GDP expenditures provided by local governments and the NBS. On the expenditure side, there are substantial differences after 2003 in investment ("gross fixed-capital formation") and net exports reported by the two sources. The sum of local investment was close to the national level until 2002. After that date, the sum of local investment exceeded aggregate investment. In 2016, the gap in the two measures of investment reached about 13 percent of GDP. The mirror image is the growing discrepancy between the sum of local net exports and aggregate net exports This gap reached about –8 percent of GDP in 2016. In contrast, the national and local differences in final consumption and changes in inventory were essentially zero after the mid-2000s.[2]

2. Final consumption includes urban and rural household consumption and government consumption. The sum of each of the local consumption components is very close to its national counterpart.

Figure 2. The Gap between Local and Aggregate GDP, by Type of Expenditure, 1992–2017[a]



Source: National Bureau of Statistics of China.
a. This figure plots the difference between the sum of provincial expenditures and aggregate expenditures as a percentage of aggregate expenditures.

To summarize the main findings: First, the sum of provincial GDP is about 5 percent higher than national GDP after the mid-2000s. Second, after 2003 the NBS adjusts industrial GDP and investment downward and adjusts net exports reported by local governments upward. Third, the NBS does not adjust local consumption—the sum of local consumption is roughly the same as the data on national consumption provided by the NBS.

I.B. Production GDP

We do not know whether the adjustments to local GDP by the NBS are appropriate. To answer this question, we need to delve into the details of the data used by the local statistical offices and the data sources behind the adjustments that are made.

INDUSTRIAL GDP Remember that the gap between the local and aggregate numbers on the production side is entirely driven by the industrial sector. The backbone source for the industry data is the Annual Survey of Industrial Firms (ASIF). These data are from a census of state-owned firms and privately owned firms with sales above 5 million yuan (until 2011) or 20 million yuan (after 2011). The Chinese statistical system calls the firms covered by the ASIF "above-scale" firms. Local statistical bureaus then add to the data from the ASIF an estimate of value added by industrial firms with sales below 5 million yuan (20 million yuan after 2011), which are referred to as "below-scale" firms in the Chinese statistical system, and businesses of self-employed individuals.[3]

We first investigate the data on value added in the ASIF. The micro data from this survey before 2007 have been widely used by researchers. After this date, however, the NBS clamped down on access to the micro data. There are good reasons to believe that the accuracy of this survey has declined over time. First, we can compare the sum of value added in the ASIF with aggregate industrial GDP reported by the NBS. This is shown in the solid line labeled "raw data" in figure 3. Aggregate value added in the ASIF should be lower than aggregate industrial GDP because the latter also includes output by small firms ("below-scale" firms) and the self-employed. However, the sum of value added in the survey exceeds the aggregate industrial GDP reported by the NBS in 2007. So the NBS must have adjusted value added in the ASIF downward.

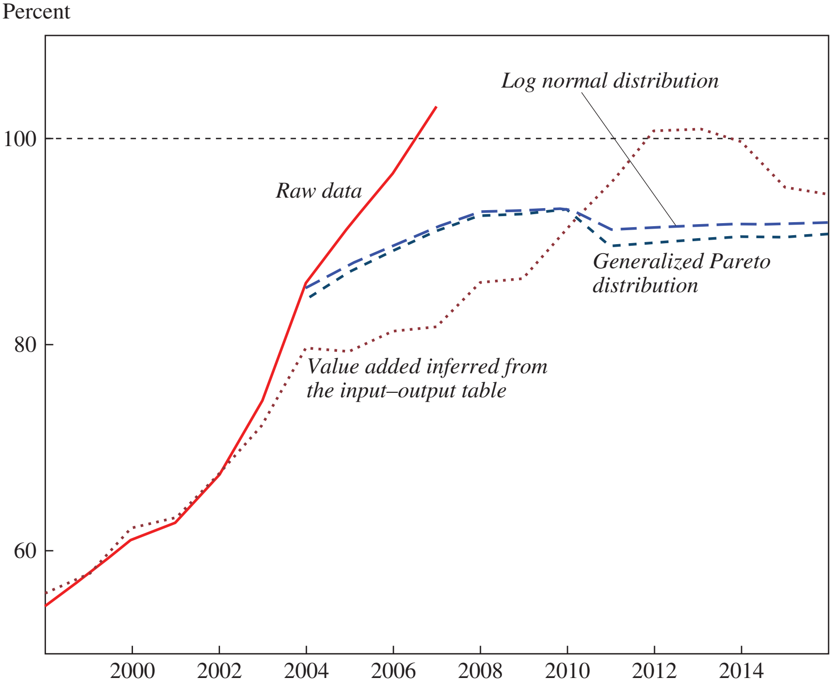
The ASIF does not report firm value added after 2008, so after this date local statistical bureaus used data on gross output in the survey to impute value added.[4] We do the same, using the ratio of gross output to value added given in the input–output (IO) tables.[5] Figure 3 presents aggregate value added imputed in this way from micro data on firm

3. The ASIF was conducted by local statistical bureaus until 2012. Orlik (2014) documents that a more centralized system was implemented nationwide in 2012, whereby firms would enter the statistics directly into an online database controlled by the NBS. Although the goal of this direct reporting system was to prevent local statistical officers from manipulating the data, local governments can still find ways to skew the data. See the case of data manipulation reported by Gao (2016) that is well known by the NBS.

4. For above-scale industrial firms and wholesale and retail firms below, we use their total sales revenue from the *China Statistical Yearbooks* to proxy total gross output.

5. The IO tables for 2002, 2007, and 2012 are from NBS (2006, 2009, 2015). The data for 2005, 2010, and 2015 are from the NBS website. For the years without data, the ratio is calculated by linear interpolation. In 2007, for example, the value-added share in industrial gross output is 0.23 and 0.29 in the IO table and ASIF, respectively. See figure A1 in the online appendix for the value-added shares between 2002 and 2015. The online appendixes for this and all other papers in this volume may be found at the *Brookings Papers* web page, www.brookings.edu/bpea, under "Past BPEA Editions."

Figure 3. Industrial GDP: The Aggregate of Micro Data versus the National Accounts, 1998–2016[a]



Sources: Authors' calculations, based on micro data from China's Industrial Survey, micro data from China's economic censuses, published tabulations from the *Input-Output Tables of China,* and data provided by the National Bureau of Statistics of China.

a. "Raw data" are the ratio of total value added by above-scale industrial firms in the Annual Survey of Industrial Firms (ASIF) micro data to industrial GDP. "Value added inferred from the input–output table" is the total value added inferred from total sales of above-scale industrial firms in the ASIF micro data adjusted by the ratio of value added to gross output in the input–output table. "Log-normal distribution" and "generalized Pareto distribution" are the total value added of above-scale firms inferred from fitting log-normal and Pareto distributions to the micro data (as a share of industrial GDP in the national accounts).

sales in the ASIF as a share of industrial GDP reported in the national accounts. The share exceeded 100 percent in 2012 and 2013. Again, the only explanation for this is that the NBS adjusted firm sales in the ASIF downward.

Remember that the ASIF only provides information for above-scale firms. For below-scale firms and the self-employed, the local statistical bureaus and the NBS rely on a survey of these two types of establishments (Xu 2004). However, the micro data from this survey are not publicly

Table 1. Aggregates in Census Micro Data versus National Accounts (percent)[a]

Year	*Above-scale firms in the census/ national accounts*	*Below-scale firms in the census/ national accounts*
Industrial firms		
2004	79.9	7.5
2008	88.1	5.6
Wholesale and retail trade firms		
2008	63.2	12.8

Sources: Authors' calculations, based on micro data from China's economic censuses and data provided by the National Bureau of Statistics of China.

a. This table reports total value added of above-scale (sales above 5 million yuan) and below-scale (sales below 5 million yuan) firms in the industrial sector (upper panel) and wholesale and retail sectors (bottom panel) as a percentage of the corresponding sectoral GDP.

available; nor is there information about the sampling and how aggregates are constructed from the survey.

We therefore take two approaches to measure the aggregate value added of small industrial firms and the self-employed. First, we use the micro data from the 2004 and 2008 economic censuses. These two censuses are a complete enumeration of all Chinese firms (including small ones), with the exception of the self-employed. The left column of table 1 shows that total value added of above- and below-scale firms in the micro data from the 2004 Economic Census is about 80 percent and 7 percent, respectively, of aggregate industrial GDP reported in the national accounts.[6] So if the 2004 national accounts are accurate, about 13 percent of industrial GDP in the national accounts is not in the census and should be attributed to the self-employed. The equivalent numbers for the 2008 Economic Census are about 88 percent and 6 percent of industrial GDP in the 2008 national accounts. The sharp increase in the output share of above-scale firms between 2004 and 2008 is consistent with the fast-growing economy, where a larger share of firms exceeds the threshold of 5 million yuan sales over time.

However, what is remarkable is that the increase in the share of above-scale firms between the 2004 and 2008 censuses reverses after 2008. The line labeled "Value added inferred from the input–output table" in figure 3 shows that the output share of above-scale firms fell by about

6. Instead of using self-reported firm value added (because of the fear that value added is inflated in the censuses, as it is in ASIF), we convert firm sales into value added by the ratio of value added to gross output in the IO tables.

6 percentage points from 2013 to 2016. If the national accounts data are accurate, the share of below-scale and the self-employed implied by the NBS's number for industrial GDP must have increased in recent years. These are precisely the firms for which the micro data are not available to the public, and this is also at odds with the trend whereby the share of above-scale firms rises over time with a growing industrial sector.

We can also estimate the importance of below-scale firms by making distributional assumptions. Specifically, we assume that firm sales follow either a log-normal or a Pareto distribution and estimate the parameters of the two distributions from the micro data from the economic census.[7] Because the economic census does not cover the self-employed, we assume that the value-added share of the self-employed in aggregate industrial GDP is about 13 percent in 2004 and about 6 percent in 2008 (see table 1). The share is linearly interpolated between 2004 and 2008 and is set to 6 percent for the post-2008 period.

Figure 3 shows the share of above-scale firms based on these two distributional assumptions. There are two main differences between the official and estimated output shares of above-scale industrial firms. First, the adjustment of the sales threshold in 2011 should generate a drop in the output share of above-scale firms in our estimates. However, there is no such drop in the official numbers.[8] Second, our estimates suggest a modest increase in the value-added share of above-scale firms since the sales threshold adjustment.[9] In contrast, the share declined after 2013 in the official numbers, which does not seem plausible.

Another way to gauge the accuracy of the NBS's estimate of industrial GDP growth is to use information on the growth of revenue from

7. We fit the two distributions by choosing parameters to fit the mean of log sales in each size percentile of industrial firms in the 2008 economic census. Specifically, we estimate the mean and standard deviation of the log-normal distribution and the mean and shape parameter of the generalized Pareto distribution to match the firm size distribution. We then assume the distribution in other years has the same standard deviation (for the normal distribution) or shape parameter (for the Pareto distribution) but a different mean parameter. We calibrate the mean parameters in the other years by targeting the average sales of above-scale industrial firms in each year. Applying the threshold of 5 million and 20 million yuan for sales before and after 2011, respectively, we can infer the output shares for above- and below-scale industrial firms.

8. In the online appendix, we present evidence that the 2010 ASIF covers fewer above-scale firms than it should. In other words, firm sales data are likely to be manipulated, disguising the otherwise discontinuous sales proportions of above-scale firms.

9. If we assume that the value-added share of individual businesses fell after 2008, as it did in the 2004–8 period, the increase in the estimated value-added share of above-scale firms would be more pronounced.

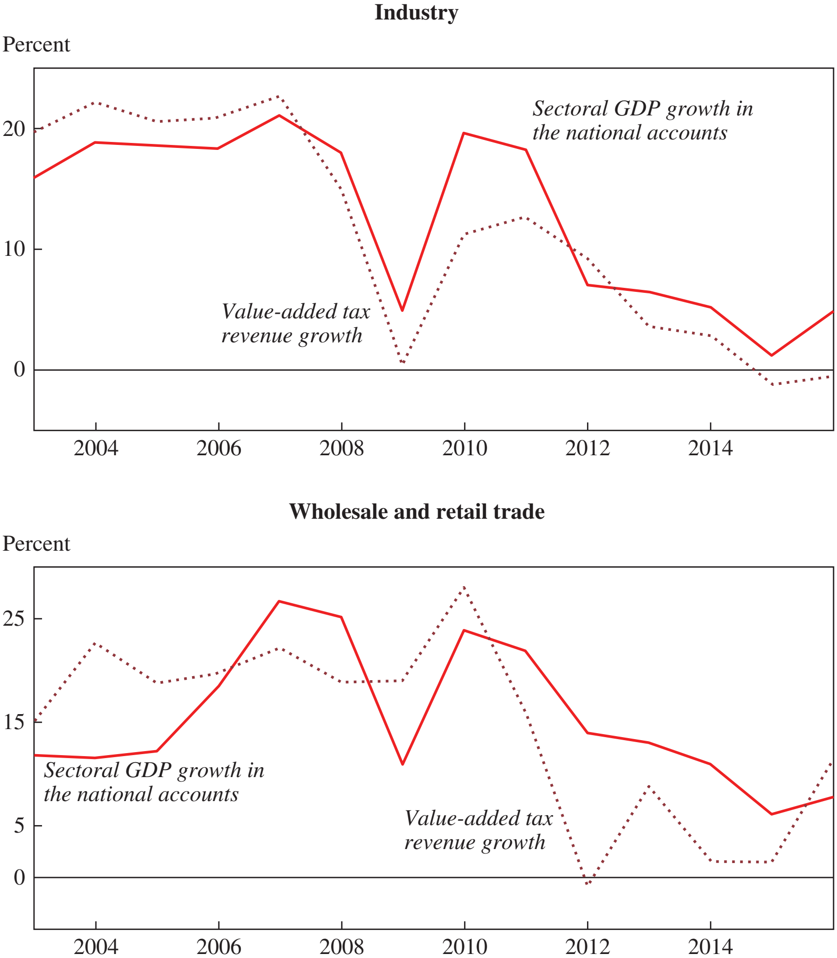
value-added taxes on industrial firms. China imposed a 17 percent value-added tax on essentially all industrial firms until 2018. There were three main exceptions. First, the value-added tax imposed on small firms (with annual sales below 0.5 million yuan) was 3 percent of their sales. Second, the value-added tax rate was 13 percent for a selected set of industrial goods. The online appendix shows that these different tax rates had negligible effects on value-added tax revenue growth. Third, a significant proportion of the domestic value-added tax (41 percent in 2015) was refundable through export tax rebates, which varied considerably across goods and over time.[10] To ensure that our estimates are not affected by tax rebates, we use data on revenues from value-added taxes *gross* of rebates for exports.

Furthermore, after 1994, there was little fraud vis-à-vis and evasion of the value-added tax. The State Administration of Taxation implemented the so-called Golden Taxation Project in 1994. A computerized taxation data network has also been in full operation since 2005, which allows the tax authorities to cross-check the input and output value-added tax at each stage of the production and distribution of goods and services. The effective value-added tax rate, defined as the ratio of the industrial value-added tax to industrial GDP net of the value-added tax, increased from 10.4 percent in 2001 to 12.9 percent in 2007, reflecting the improved tax enforcement in the period. There are several possible reasons why the effective tax rate was below the main statutory tax rate of 17 percent. The different tax rates mentioned above matter, but they cannot be quantitatively important. Another possibility is inflated industrial output. To account for this difference of nearly 4 percentage points, industrial GDP would need to be overestimated by nearly a quarter in 2007. Although tax evasion is hard to do for transactions within the industrial sector, it may be easier for industrial output sold to the sectors to which value-added tax does not apply.[11] But even if there is some tax fraud and evasion, as long as their degree does not increase, revenues from the value-added tax on industrial firms should be proportional to industrial GDP.

10. The domestic value-added tax revenue and the tax rebate for exports can be found in the *China Taxation Yearbook*.

11. The 2007 IO tables suggest that the share of industrial output to the construction and service sectors (excluding wholesale and retail) is 15.9 percent. If industrial firms hide all their output sold to the construction and service firms that do not have incentives to ask for value-added tax invoice, the effective value-added tax rate would be lowered by 2.7 percentage points.

Figure 4. The Growth in Value-Added Tax Revenues and GDP, 2003–16



Sources: National Bureau of Statistics of China; *China Taxation Yearbook*.

The top panel of figure 4 compares the growth rate of revenues from domestic value-added taxes with the growth rate of industrial GDP. The growth rate of revenues from value-added taxes exceeds that of industrial GDP before the mid-2000s, consistent with the improved enforcement of value-added taxes. However, after 2007, the growth rate of tax revenues is lower than the growth rate of industrial GDP. Furthermore, the gap has been widening over time. In 2010 to 2012, for instance, value-added tax revenue growth is about two-thirds that of industrial GDP. The growth in tax revenues dropped to about half the growth rate of industrial GDP growth in 2013 and 2014, and even became negative in 2015 and 2016. Consequently, the effective value-added tax rate fell from about 12.9 percent in 2007 to about 9.3 percent in 2016.

A few tax policies introduced after 2007 may lower the effective value-added tax rate. The most relevant change is the value-added tax deduction on fixed-asset investment for domestic firms.[12] This policy was first introduced in three provinces of Northeast China in 2004, and was later extended to six provinces (Cai and Harrison 2018; Zhang, Chen, and He 2018). The central government unexpectedly increased the coverage to all provinces at the end of 2008, as part of the stimulus package in response to the global financial crisis. The nationwide policy became effective on January 1, 2009. Although the policy obviously reduced industrial value-added tax revenue in the transition period between 2004 and 2009, it is hard to estimate the extent to which value-added tax revenue growth was affected. The main obstacle is that we do not know how much of fixed-asset investment is deductible from the value-added tax.[13] A simple fix is to look at the value-added tax revenue growth after 2009.[14] The average industrial value-added tax revenue growth was about 5.3 percent between 2009 and 2016, which is about 3.4 percentage points lower than the average industrial GDP growth in the same period. The gap is similar to that of about 3.5 percentage points in the period 2007–16.

Another important policy change is the reform of replacing business tax with value added tax initiated in 2012 and completed in 2016. Because the purchase of service goods became deductible from value-added tax,

12. Foreign firms have always been eligible for the tax deduction.

13. The value-added tax deduction only applies to purchase of machinery, mechanical apparatus, means of transportation and other equipment, tools and fixtures related to production, and business operations (Ministry of Finance and State Administration of Taxation 2008a). According to the compositions in China's fixed-asset investment survey, which severely overestimated the level of fixed-asset investment, as is shown below, purchase of equipment and instruments accounts for about 40 percent of fixed-asset investment in the industrial sector. But the purchase of equipment and instruments includes items such as those for the newly built production department that are not eligible for value-added tax deduction.

14. The remaining concern is that the effect of the policy might be persistent by increasing industrial firms' investment rate in the subsequent periods. Using firm survey data from China's State Administration of Taxation, Chen and others (2019) found that the average firm investment rate increased by 2.6 percentage points in 2009 but then decreased by 0.7 percentage point in both 2010 and 2011. The diminishing effect on investment rate suggests that the policy should not lower value-added tax revenue growth in 2010 and onward.

the reform may also lower value-added tax revenue growth in the industrial sector. The pilot started in the transportation industry (excluding railway) and several "modern" service sectors in Shanghai from January 2012. The policy was extended to Beijing and seven other provinces and cities from August 2012 and then to all provinces from August 2013.[15] Railway transportation, the postal service, and telecommunications were added to the list of "modern" service sectors in 2014. In the online appendix, we identify 10 industries in the IO tables according to the description of "modern" service sectors in the documents. The transportation industry and all the modern service sectors accounted for 2.4 percent and 2.3 percent of industrial input, respectively, in 2012 (see the online appendix). For two main reasons, we assume that this reform did not affect industrial value-added tax revenue growth between 2010 and 2016. First, the purchase of transportation services was deductible from the value-added tax even before the reform started. Second, the input share of the modern service sectors was not big enough to generate a significant effect on industrial value-added tax revenue growth in the reform period.

Finally, the value-added tax policy for small taxpayers (defined as those with annual sales below a half million yuan) was adjusted twice after 2007. From August 2013 onward, taxpayers with annual sales below 240,000 yuan were exempted from the value-added tax. The cutoff was increased to 360,000 yuan in October 2014.[16] The effectiveness of the tax reform can be seen from the share of the value-added tax paid by small taxpayers, which is publicly available in the *China Taxation Yearbook*. In fact, the share was quite stable, at about 4 percent, between 2010 and 2016, and even increased from 3.7 percent in 2012 to 4.4 percent in 2013. One explanation is that firms and individual businesses with annual sales below the cutoff contribute little to total value-added tax revenue.

To summarize the main findings about the reliability of the NBS's estimate of industrial GDP: First, the micro data from the ASIF have overstated aggregate output at least since 2007. Second, the aggregate industrial GDP provided by the NBS implies an increasing share of below-scale firms and the self-employed in the industrial sector after 2012. Third, the growth rate of the aggregate industrial GDP has exceeded the growth rate of revenues from value-added taxes on industrial firms since 2008. Based on these three pieces of evidence, we conclude that despite the

15. See Ministry of Finance and State Administration of Taxation (2011).

16. See Ministry of Finance and State Administration of Taxation (2013, 2014). See also Lardy (2014) for the evolution of policies toward the private sector.

adjustments made by the NBS to local industrial GDP, the official numbers for the aggregate industrial GDP—and by extension, the aggregate GDP for all sectors—are likely to overstate the truth after 2007–8.

NONINDUSTRIAL GDP Turning to the nonindustrial sector, the NBS conducts surveys for all "qualified" construction firms, above-scale wholesale and retail firms, above-scale hotel and catering firms, and all real estate developers and operators.[17] We first look into the wholesale and retail trade sectors, which accounted for about 10 percent of the aggregate GDP in 2016. Though the published tabulations of the surveys provide total sales of above-scale wholesale and retail firms, value added is not reported. We thus convert total sales to value added, following the procedure used by Bai and others (2019).

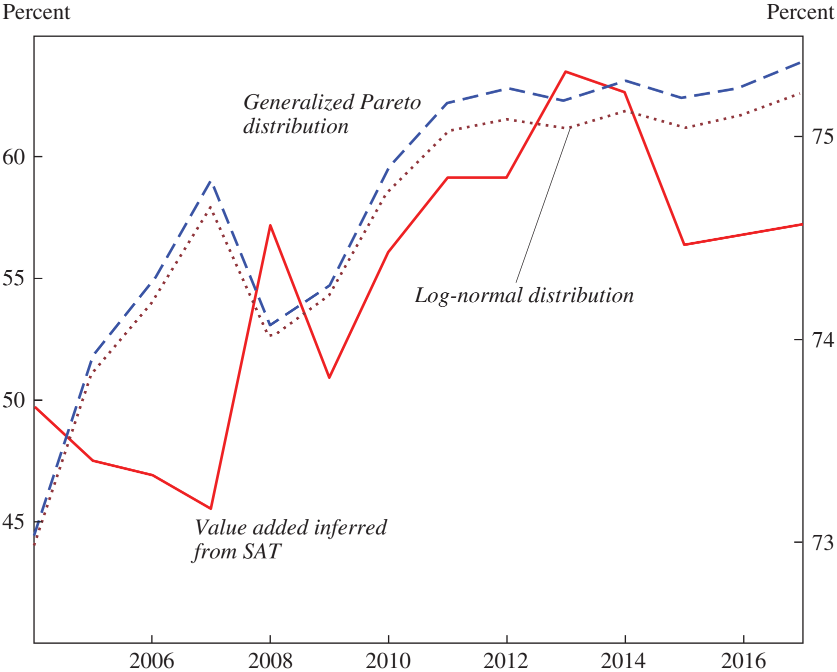
Based on this imputation, the solid line in figure 5 plots the value added of above-scale wholesale and retail firms in the published surveys as a share of official aggregate GDP in the wholesale and retail trade sectors. Using the same procedure described in the previous section, we estimate the parameters of the distribution of firm size in the wholesale and retail sectors in the 2008 Economic Census. We then calibrate the mean parameters of the log-normal and Pareto distributions to match average firm sales in each year. We further assume the value-added share of the self-employed in wholesale and retail GDP is fixed at 24 percent (the number suggested by the 2008 Economic Census; see table 1). The estimated models suggest that the share of above-scale wholesale and retail firms in aggregate GDP in these sectors has increased slightly in recent years. Like what we see for the industrial sector, this is also at odds with the dramatic drop in 2014 and 2015 in the official data.

The bottom panel of figure 4 compares domestic value-added tax revenue growth from the wholesale and retail sector with GDP growth in these sectors as provided by the national accounts.[18] Like what happened in the industrial sector, tax revenue outgrew sectoral GDP before the mid-2000s, but the pattern was reversed after 2010, except for 2016. The average difference from 2010 to 2016 between tax revenue and GDP growth is about 6 percentage points, suggesting that true wholesale and retail GDP is also likely to be overstated in the national accounts.

17. The sales threshold for wholesale and retail firms is 20 and 5 million yuan, respectively. The sales threshold for hotel and catering firms is 2 million yuan.

18. Value-added tax revenue accounts for about 40 percent of total tax revenue in the industrial and the wholesale and retail trade sectors. See figure A4 in the online appendix.

Figure 5. Wholesale and Retail GDP: The Aggregate of Micro Data versus the National Accounts, 2004–17[a]



Sources: Authors' calculations, based on micro data from China's economic censuses, published tabulations from the *Input-Output Tables of China*, and data provided by the National Bureau of Statistics of China.

a. This figure presents the ratio of the value added of above-scale retail and wholesale firms to wholesale and retail GDP in the national accounts. "Value added inferred from SAT" is the value added imputed from the total sales of above-scale wholesale and retail firms in the firm survey adjusted by the ratio of value added to sales in the data from the State Administration of Taxation (SAT). "Log-normal distribution" and "generalized Pareto distribution" are the total value added of above-scale firms inferred from fitting log-normal and Pareto distributions to the micro data.

The construction sector accounts for about 7 percent of GDP after 2010 (see the bottom panel of figure 1). Surprisingly, the output share of "qualified" construction firms has fallen in recent years (see figure A2 in the online appendix). Although no sales threshold applies to construction firms, larger construction firms are more likely to be qualified. For the same reason discussed above, pure economic forces are hard to reconcile with the observed output share change.

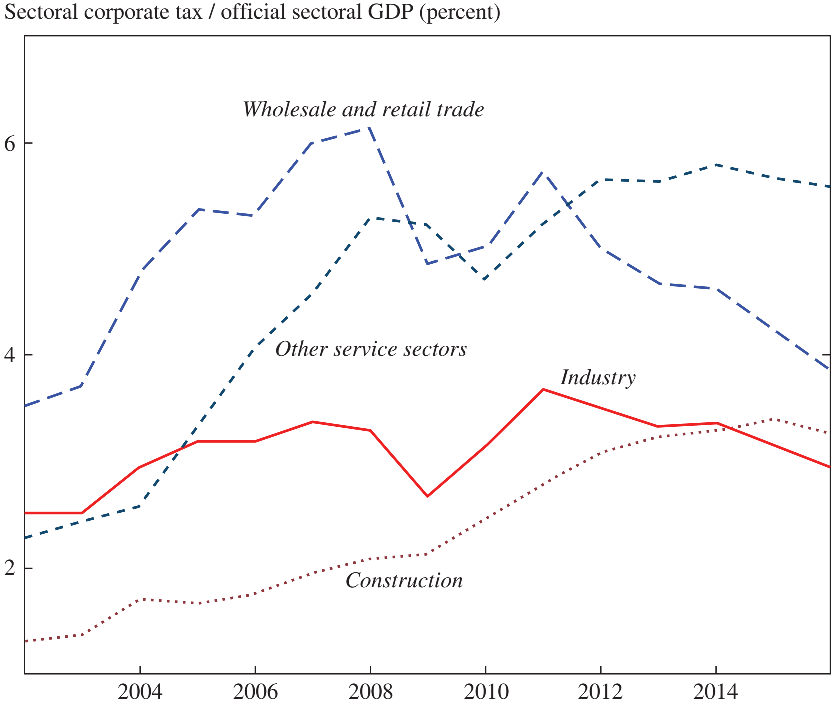
It is more difficult to examine the reliability of nonindustrial GDP because we do not have access to firm-level data other than the 2008

Economic Census. Furthermore, the value-added tax only applied to the industrial and wholesale and retail sectors before 2017. A close substitute is the corporate income tax. Like the value-added tax, a major proportion of corporate income tax revenue (60 percent) is paid to the central government.[19] Unlike the highly rigid value-added tax rate, there are many exemptions and special rates for corporate taxes. For example, there are special corporate income tax rates for labor-intensive and high-technology firms. The enforcement of corporate income taxes is also weaker than that of value-added taxes. Figure 6 plots sectoral corporate income tax revenue as a percentage of sectoral GDP for these four sectors: industry, construction, wholesale and retail trade, and services excluding wholesale and retail trade. The ratio of corporate income tax revenue to GDP increased in all the sectors before 2007. This is likely to be driven by both growing firm profitability and enhanced tax enforcement in the period. The ratio of corporate income tax revenue to GDP decreased dramatically in industry and in wholesale and retail trade after 2011, consistent with the growth slowdown in the two sectors. The ratio was fairly stable in the service sector excluding wholesale and retail, by about 2 or 3 percentage points higher than the ratio in industry and wholesale and retail trade in recent years. Construction is the only sector where the ratio of corporate income tax revenue to GDP kept increasing until 2015.

That said, corporate income tax revenue is still informative. Figure 7 compares corporate income tax revenue growth with GDP growth in the four sectors. For industry and wholesale and retail trade (respectively, the figure's top left and bottom left panels), the results are similar to those in figure 4: The sectoral GDP growth is above tax revenue growth in recent years. For construction (the top right panel), corporate income tax revenue growth is above GDP growth in most years. Given the fact that the ratio of corporate income tax revenue to GDP was very low in the construction sector in earlier years (figure 6), the strong corporate income tax revenue growth might be a consequence of much improved tax enforcement in that sector. Most interestingly, the bottom right panel of figure 7 shows that GDP growth seems in line with corporate income tax revenue growth in

19. The corporate income tax revenue paid to the central/local government is from the *China Taxation Yearbook*. A total of 75 percent and 50 percent of value-added tax revenue was paid to the central government before and after 2016. The sharing mechanism prevents local governments from inflating corporate income tax revenue, which would otherwise incur direct losses to local fiscal budgets. There is evidence that local governments manipulate business tax revenue, which applied to most service sectors and went entirely to local fiscal budgets (Lei 2017). The business tax was replaced by the value-added tax in 2017.

Figure 6. Corporate Income Tax Revenue as a Percentage Share of GDP, 2002–16[a]



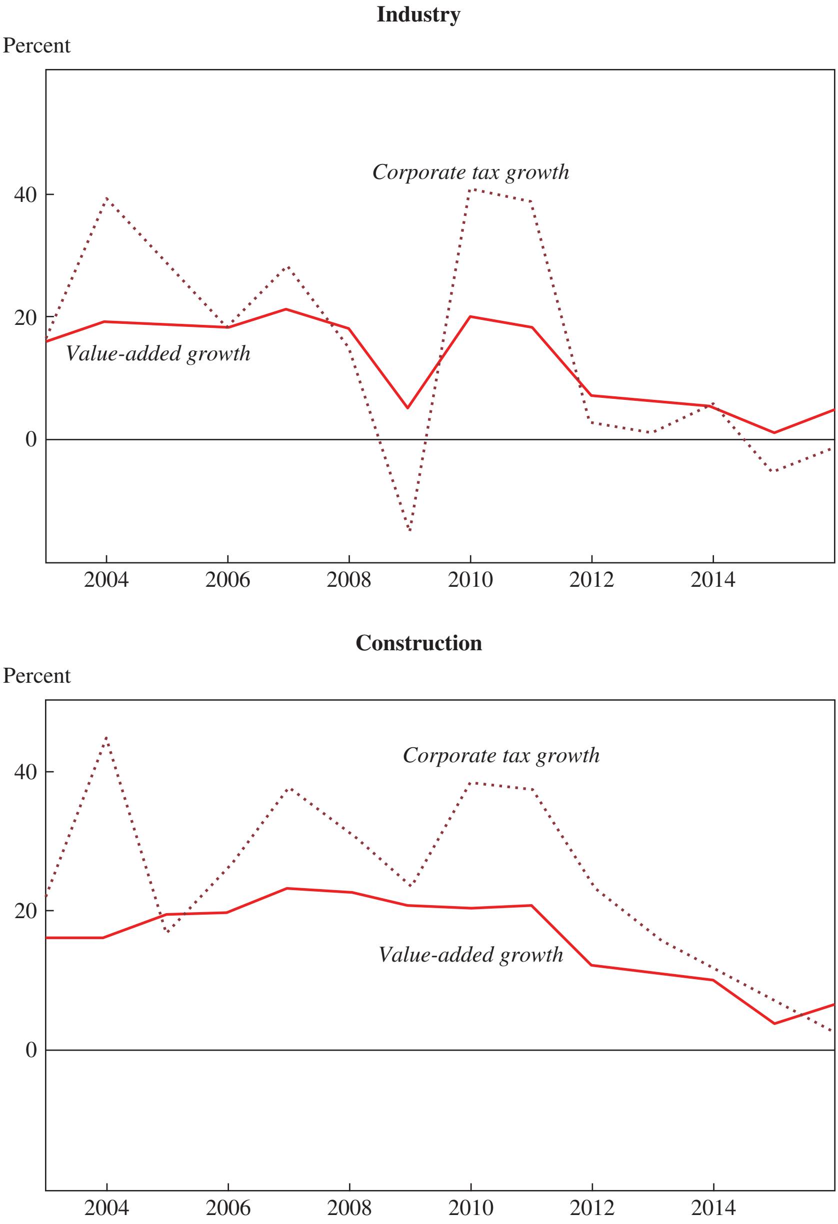
Sources: National Bureau of Statistics of China; *China Taxation Yearbook.*

a. This figure plots the ratio of corporate income tax revenue to GDP in the industrial sector, construction, wholesale and retail trade, and other services excluding wholesale and retail.

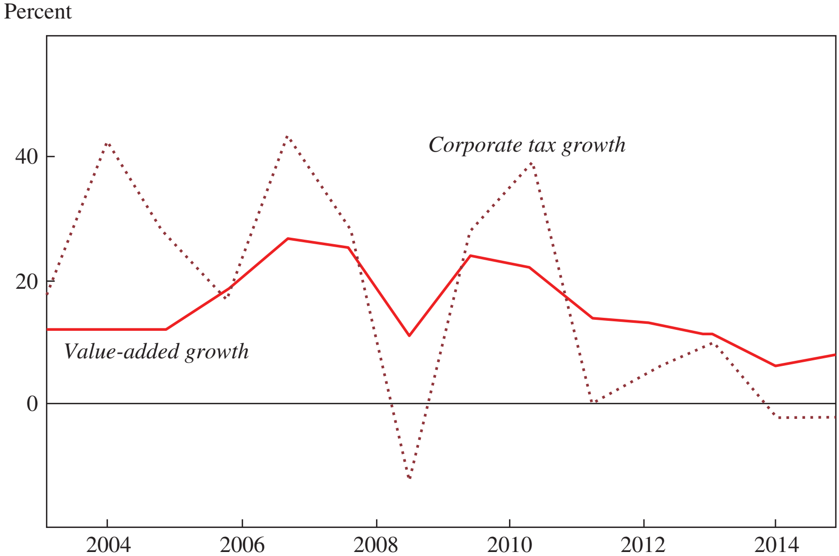
the service sector excluding wholesale and retail trade. Because the ratio of corporate income tax revenue to GDP did not change much, we view the figure's bottom right panel as evidence that official estimates of GDP growth in the service sector (excluding wholesale and retail trade) in the national accounts are reliable.

In sum, while the growth in the wholesale and retail trade sectors is likely to be overstated in the official statistics, there is no evidence that the official statistics for the other service sectors are inaccurate. However, the effect of inaccuracies in the wholesale and retail trade sectors is important, given that these are two large sectors. Note also that figure 2 shows no gap between local and aggregate statistics for the service sector after 2003. Figure 2 simply tells us exactly where the NBS has adjusted the local numbers, not whether the adjustment or the absence of an adjustment is appropriate.

Figure 7. Sectoral GDP and Corporate Income Tax Revenue Growth, 2003–16



Wholesale and retail trade



Other service sectors

Sources: National Bureau of Statistics of China; *China Taxation Yearbook.*

I.C. Expenditure GDP

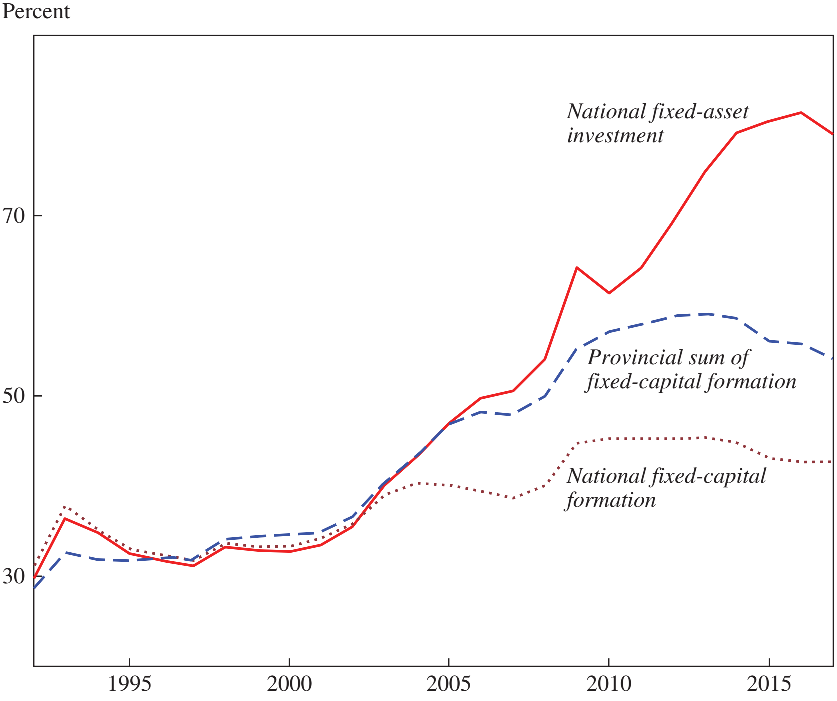
We now examine the underlying data used to construct GDP expenditures. As discussed above, government expenditures reported by local governments are consistent with those reported by the NBS. Furthermore, this information is based on administrative and verifiable data on public expenditures, so it is likely to be reliable. We therefore focus on household consumption, investment, and net exports.

The urban and rural household surveys are the backbone of aggregate household consumption. From these two surveys, the local statistical bureaus and the NBS directly take aggregates of household spending on food, clothing, household facilities, education, culture and recreation services, and miscellaneous goods and services. The other components of household consumption also use the household surveys but are adjusted for (1) accounting discrepancies (that is, medical expenditures paid by the government are not in the two household surveys but are included in final consumption); (2) biases in the surveys (that is, high-income households are underrepresented in the two household surveys). Xianchun Xu (2014) describes in detail how the NBS arrives at consumption aggregates by adjusting the data from the two household surveys. These adjustments are based on administrative data from the relevant government departments. For instance, the NBS uses social security income and expenditure data to adjust medical expenditures. Another example is to use the production, sales, and import data on automobiles from the Association of the Automobile Industry and the Department of Public Security (which registers all new automobiles) to adjust consumption data for transportation and communication. This helps correct the bias caused by underrepresented high-income households, which are more likely to purchase automobiles.

Investment spending is officially called fixed-capital formation (FCF) in the Chinese national accounts. These data are primarily based on reports of fixed-asset investment (FAI) by local governments. FAI measures gross investment spending, given that it includes expenditures on land purchases and used capital. Therefore, local statistical authorities use a survey of land purchases and used capital to subtract these two items from FAI in order to estimate net investment spending—that is, FCF.

However, there is abundant evidence that local data on gross investment have become more unreliable. In contrast with ASIF, which is based on a firm's financial statement, local administrative data on gross investment are based on reports of investment projects by local governments. There is no audit of these data; nor are there any consequences for misreporting

Figure 8. Fixed-Asset Investment versus Fixed-Capital Formation, 1992–2017



Source: National Bureau of Statistics of China.

this information. In addition to the incentives for local officials to misreport this number, tax considerations may also lead to the inflation of FAI.[20] In 2014, Xu Xianchun, a vice director of the NBS at the time, publicly stated that FAI is inflated by local statistical offices (Xu 2014). According to him, "Some regions tend to set unrealistic investment program targets, which are in turn assigned to lower-level governments as performance indicators."

Figure 8 shows that the gap between FAI and national FCF has increased since the early 2000s.[21] In 2015, the gap between aggregate FAI and net investment provided by the NBS (that is, FCF) reached about 38 percent

20. For instance, the Ministry of Finance and the State Administration of Taxation (2008b) introduced a policy that allows taxpayers to deduct fixed-asset investments from the value-added tax.

21. National and provincial FAI data are from the *China Statistical Yearbook*. The national FAI is identical to the sum of provincial FAIs.

of official GDP. In theory, the gap between the two measures of investment should only reflect land purchases and spending on used capital. The purchase of land and used capital does account for most of the difference in the early 2000s, but these two items are much too small to account for the gap in recent years (see figure A3 in the online appendix).[22] The enormous gap between FAI and FCF suggests that investment spending is overstated by local statistical offices and that the NBS has made large adjustments to these data to arrive at a number for aggregate net investment.

It is also evident that even local statistical bureaus adjust FAI downward when estimating local net investment. Figure 8 shows that the sum of provincial FAI exceeds provincial FCF by about 24 percent of GDP in 2015. This difference is, once again, too big to be reconciled by accounting discrepancies like purchase of land and used capital. But the extent to which local statistical bureaus adjust the data on FAI is obviously less than the adjustment by the NBS. The sum of FCF at the provincial level exceeds aggregate FCF by about 14 percent of GDP in 2015.

Note that the adjustment made by the NBS to investment spending provided by the local statistical bureaus is larger than the adjustment made to local estimates of industrial GDP. Because local GDP on the production side has to be equal to GDP on the expenditure side, local statistical bureaus use local net exports as the residual to balance production and expenditure GDP. This can be seen in figure 2, where the growing discrepancy between net exports and local net outflows is the mirror image of the gap between national and local FCF in figure 8.

The NBS completely disregards local estimates of net exports. Instead, it calculates aggregate net exports from data on net exports of goods in the customs data. For this reason, aggregate net exports in the national accounts are very close to net exports reported in the customs data. In contrast, local estimates of net exports are not based on any data and are simply a residual used to equalize local production and expenditure GDP.

To summarize the main findings: Local statistical bureaus inflate investment and, to a smaller extent, inflate output in the industrial and wholesale and retail trade sectors. Because investment data are easier to manipulate

22. Holz (2013, 2018) also documents the growing discrepancies between provincial and national investment and the widening gap between FAI and FCF. Liu, Zhang, and Zhu (2016) also show that the gap between FAI and FCF cannot be explained by land sales and purchases of used assets and buildings. Data for land sales and purchases of used assets and buildings are from the *China Land and Resources Statistical Yearbook* and the *Statistical Yearbook of the Chinese Investment in Fixed Assets*, respectively.

(the amount of investment is project-specific, and is disconnected from investing firms' financial statements), the misstatement of investment spending is more severe than the bias in GDP. The gap between the two is "reconciled" by the large net inflows of goods and services reported by local governments. In contrast, consumption data based on household surveys are more reliable.

II. Revised Estimates of GDP Growth

Given the findings presented above, this is the obvious question: What are the "true" estimates of China's GDP growth? Here, we make two efforts to come up with a number. First, we use alternative data from tax records to generate alternative measures of GDP on the production side. We then use these measures to reestimate aggregate investment as well as local GDP. Second, we take a data-fitting approach and use external data that are not likely to be manipulated by local governments to estimate GDP.

II.A. Adjusting National Accounts with Tax Data

Our first approach to estimate "true" GDP is built on three assumptions. First, we assume that industrial output reported by local statistical officers has not been reliable since the late 2000s. Second, we assume that nonindustrial output reported by local statistical officers is reliable. And third, we assume that industrial value-added tax revenue is proportional to true industrial value added.

The validity of the first assumption comes from the facts in the previous section. In particular, industry is the only major sector for which the NBS significantly adjusts locally reported output data. The second assumption is partly based on the evidence that corporate income tax revenue grew in tandem with value-added taxes in the service sector, and is partly made for practical reasons, given that we do not have reliable data to indicate true output in most nonindustrial sectors.[23] We relax the second assumption below. The third assumption is the strongest one, because it hinges on two institutional features discussed in the previous section. First, China has developed a sophisticated value-added taxation system to minimize tax fraud and evasion. Second, local governments do not

23. See also Bai and others (2019) for more evidence on the reliability of service data in the national accounts.

have incentives to overstate value-added tax revenue, because otherwise they would incur direct local fiscal losses. We have also discussed several tax policy changes that are likely to impair the proportionality between industrial GDP and value-added tax revenue. The effect of replacing the business tax is quantitatively small before 2016, and so is the effect of value-added tax exemption on small firms and individual businesses. Our adjustment begins with 2010, after the end of the value-added tax deduction on fixed-asset investment.

In the simplest case, our adjusted GDP assumes this equation:

$$\text{(1)} \qquad \text{Adjusted GDP}_t = \text{Official GDP}_t - \Delta \text{ Industrial GDP}_t$$

where $\Delta X_t \equiv$ Official X_t – Adjusted X_t denotes the adjustment in variable X and

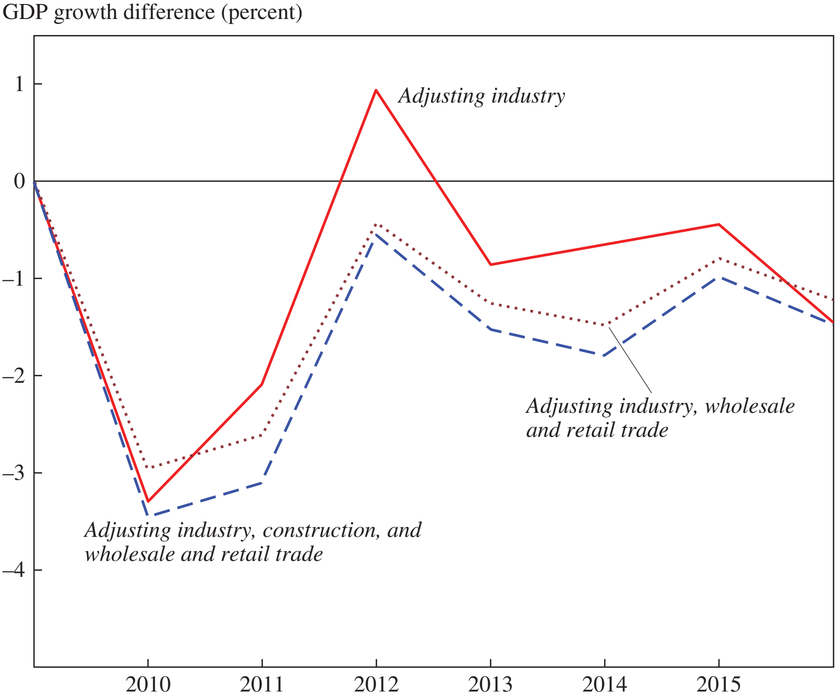
$$\text{Adjusted Industrial GDP}_t = \text{Adjusted Industrial GDP}_{t-1} \cdot \text{Industrial VA Tax Revenue Growth}_t.$$

The dotted line in figure 9 plots the difference between our adjusted GDP growth and the official nominal GDP growth (the solid line). The adjusted growth rate is always below the official growth rate, except in 2012. Figure 4 above shows that industrial value-added tax revenue growth is about 3.4 percentage points lower than official industrial GDP growth after 2009. The industrial sector accounts for roughly one-third of China's GDP. Therefore, correcting the overreporting of industrial output lowers GDP growth from 2010 to 2016 by about 1.1 percentage points; see the second column of table 2.

We relax the third assumption by also adjusting value added in wholesale and retail output growth. Because wholesale and retail value-added tax revenue growth is also below its GDP growth in the national accounts (figure 4), adjusting output in both the industrial and wholesale and retail sectors would further cut nominal GDP growth in recent years (the dashed line in figure 9). After we also adjust the growth rate of the wholesale and retail sectors, our estimate of the growth rate of nominal GDP from 2010 to 2016 is about 1.5 percentage points lower than the official rate (the third column of table 2).

We next look into expenditure-side GDP accounting. Based on the discussions in the previous section, we assume that the official statistics

Figure 9. The Gap between Adjusted and Official Nominal GDP Growth, 2009–16[a]



Sources: Authors' calculations, based on data provided by the National Bureau of Statistics of China and published tabulations from the *China Taxation Yearbook*.

a. This figure plots the difference between adjusted nominal GDP and official nominal GDP growth rates. The solid line adjusts industrial GDP growth only. The dotted line adjusts both industry and wholesale and retail trade GDP growth. The dashed line adjusts industry, construction, and wholesale and retail trade GDP.

on aggregate consumption and net exports are accurate. FCF is then obtained by

$$\begin{aligned}\text{Adjusted FCF} = {} & \text{Adjusted GDP} - \text{Final Consumption} \\ & - \text{Inventory Change} - \text{Net Exports.}\end{aligned}$$

However, this adjustment is incomplete. Most of the output of the construction sector is classified as investment on the expenditure side of GDP. Although the NBS does not adjust local estimates of the construction sector's output (the bottom panel of figure 1), our estimated investment

Table 2. The Adjusted Growth Rate of Nominal GDP[a]

		Adjustment by value-added tax			*Adjustment by statistical model on industrial GDP*		
Year	*Official data*	*Adjusting industrial GDP*	*Adjusting industrial and W&R GDP*	*Adjusting industrial, construction, and W&R GDP*	*All variables*	*Without light*	*Without light and national tax*
2009	9.25	9.25	9.25	9.25	7.09	7.97	5.26
2010	18.32	15.02	15.36	14.86	20.27	19.13	19.78
2011	18.47	16.37	15.86	15.36	19.42	18.17	17.76
2012	10.44	11.38	10.00	9.88	9.98	9.95	9.33
2013	10.16	9.29	8.89	8.63	10.12	9.51	10.08
2014	8.19	7.54	6.70	6.39		7.56	7.04
2015	7.00	6.55	6.20	6.02		5.98	4.83
2016	7.91	6.46	6.69	6.43			7.33
2017	10.93						10.19
Average, 2010–16	11.50	10.37	9.96	9.65			10.88

Sources: Authors' calculations, based on data provided by the National Bureau of Statistics of China and published tabulations from the *China Taxation Yearbook*.
a. W&R = wholesale and retail sector.

spending from the equation above suggests that the construction sector's output is also overstated. We therefore adjust the construction sector's GDP using this formula:

$$\text{(2)} \qquad \Delta \text{ Construction GDP}_t = \frac{\gamma_t}{\theta^I_{con,t}} \cdot \Delta \text{ FCF}_t,$$

where $\theta^I_{con,t}$ denotes FCF per unit of construction GDP and γ_t is the proportion of construction FCF in total FCF.[24] The adjustment in the construction GDP leads to further adjustment in the aggregate GDP and, hence, another round of adjustment in the FCF and construction GDP. The full adjustment that balances the aggregate GDP, construction GDP, and FCF is given by:

$$\text{(3)} \qquad \text{Adjusted GDP}_t = \text{Official GDP}_t + \frac{1}{1 - \gamma_t / \theta^I_{con,t}} \Delta \text{ Industrial GDP}_t.$$

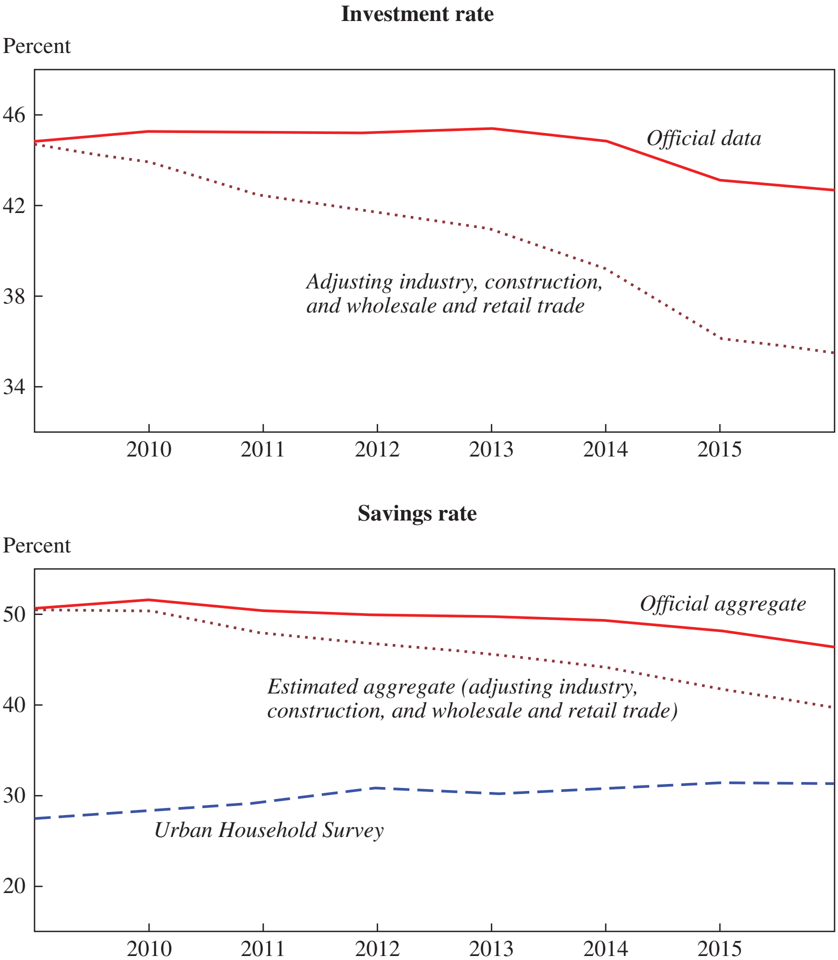
Compared with equation 1, the GDP adjustment in equation 3 is amplified by adjusting construction output. When we also adjust wholesale and retail GDP, Δ Industrial GDP_t in equation 3 should be replaced by Δ Industrial GDP_t + Δ WR GDP_t, where WR GDP denotes wholesale and retail GDP.

The results are shown in the top panel of figure 10. As can be seen, our estimate of the investment rate is significantly lower than the official numbers. In 2016, we estimate that the investment rate is about 35.6 percent of GDP—the official number is about 7 percentage points higher. Looking at the change since 2010, our estimate is that the investment rate fell from about 43.9 percent in 2010 to about 35.6 percent in 2016. The official statistic is that the investment rate decreased from about 45.2 percent to about 42.7 percent between these two years. Figure A6 in the online appendix plots the implied construction GDP growth.

The fourth column of table 2 reports the growth rate of nominal GDP after all three adjustments (industrial, wholesale and retail trade, and construction output). With all three adjustments, nominal GDP growth since 2013 has been about half the official growth rate of nominal GDP. During the 2010–16 period, our estimate of GDP growth is about 1.8 percentage points lower than the official growth rate.

24. Note that we do not need to adjust industrial output in a similar fashion. This is because industrial output can be exported, while construction output is for domestic use.

Figure 10. Official and Adjusted Investment Rates, 2009–16[a]



Sources: Authors' calculations; National Bureau of Statistics of China.

a. The solid line in the top panel plots the ratio of official fixed-capital formation (FCF) to official GDP. The dotted line plots the ratio of *adjusted* FCF and *adjusted* GDP (with industry, construction, and wholesale and retail trade GDP adjusted). For comparability with the adjusted FCF afterward, in 2009, the adjusted FCF is calculated as production-side GDP minus other expenditure components. The solid line in the bottom panel is the aggregate savings rate in the national accounts. The dotted line plots the adjusted aggregate savings rate. The dashed line is the average savings rate in the Urban Household Survey.

The bottom panel of figure 10 shows our estimate of the savings rate. Our estimate is that the savings rate fell significantly between 2010 and 2016, from about 50.4 percent to about 39.7 percent of GDP. The official numbers show a much smaller decrease, from about 51.5 percent to about 46.4 percent. Figure 10 also shows that our revised estimate of the savings rate is closer to the savings rate computed from the micro data from the Urban Household Survey. The smaller difference implies a more reasonable saving rate in the nonhousehold sector. Household income accounts for 62 percent of GDP in 2016.[25] To reconcile the official aggregate saving rate of about 46 percent and the household saving rate of about 31 percent, we would need a saving rate of about 70 percent in the corporate and government sectors. If, instead, the aggregate saving rate follows our estimate, the corporate and government sectors would have a saving rate of 54 percent, which is more reasonable than what is implied by the official aggregate saving rate.[26]

II.B. Adjusting Local GDP

A similar procedure can be applied to correct provincial GDP. The published data on revenues from value-added taxes do not break down revenues by province industries. However, value-added tax revenues from industry and wholesale and retail trade account for more than 90 percent of total value-added tax revenues before 2015 (see figure A5 in the online appendix). We use provincial value-added tax revenue growth to proxy the value-added tax revenue growth of industry and wholesale and retail trade in the provinces. The same benchmark adjustment for national GDP can then be used for provincial GDP.[27]

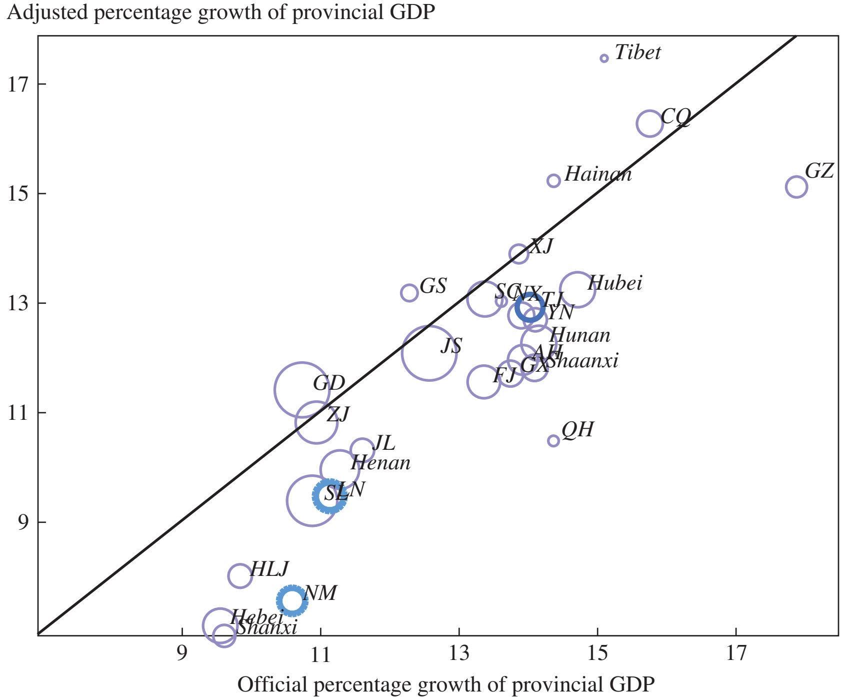
Figure 11 shows a scatterplot of our adjusted growth rate of provincial GDP against the official growth rate of provincial GDP. The majority of the provinces lie below the 45-degree line, indicating that the official growth rate of most provinces exceeds our adjusted estimates. The average

25. The data on household income and GDP are from the Flow of Funds Account in the *China Statistical Yearbook*.

26. The household saving rate may be underestimated in the surveys. Using the household saving rate of 36 percent in the 2016 Flow of Funds Accounts, the implied saving rate in the corporate and government sectors would be 62 percent and 46 percent by the official and our adjusted aggregate saving rate, respectively.

27. We drop Shanghai and Beijing for two reasons. First, these two provinces replaced the business tax with the value-added tax in 2012 and 2013, respectively. The reform had a significant effect on value-added tax revenue in both provincial-level cities, where the service sector is substantially larger than the industrial sector. Second, it is widely acknowledged that the two cities are among the regions with the most reliable GDP data.

Figure 11. Official and Adjusted Provincial GDP Growth, 2009–15[a]



Sources: Authors' calculations, based on data provided by the National Bureau of Statistics of China and published tabulations from the *China Taxation Yearbook*.

a. This is a scatterplot of the *adjusted* (*y* axis) versus official (*x* axis) provincial annualized GDP growth rate between 2009 and 2015. The three highlighted provinces are Liaoning (LN), Inner Mongolia (NM), and Tianjin (TJ), which were involved in GDP scandals. The solid line is a 45-degree line.

difference is about 1.2 percentage points. Guangdong (GD) and Zhejiang (ZJ) , however, are located on the 45-degree line. Among the provinces that are far below the 45-degree line are Liaoning (LN) and Inner Mongolia (NM). Local leaders in the two provinces were recently arrested in corruption crackdowns, and one of the official accusations was that these leaders had overstated local GDP. In addition, after the corruption crackdown, the local statistical bureaus in LN and NM issued new revised estimates of local GDP in 2016 and 2017, respectively.[28] The new numbers are

28. Tianjin (TJ) also acknowledged that the Binhai district overstated its GDP. But the Tianjin municipal government claimed that the district-level GDP overstatement did not affect Tianjin's GDP.

about 22 percent and about 11 percent lower than the official numbers in the previous year.[29] In comparison, our estimates show that the unadjusted official GDP in LN and NM was overstated by about 9 percent and about 15 percent in 2015, respectively. Furthermore, the official adjustment on industrial GDP accounts for about 70 percent of its adjustment on GDP in LN. In the case of NM, the local statistical bureau revised its estimate of total value added of above-scale industrial firms in 2016 downward by about 290 billion yuan, which accounts for the entire downward revision in GDP of NM that year.

Adjusting local FCF is more difficult. Unlike net exports at the national level, which are underpinned by customs data, provincial net outflows of goods and services are not based on any data. Therefore, the adjustment for national FCF cannot be applied to provincial FCF. We can, however, use this equation to indicate provincial FCF:

$$\text{(4)} \qquad \mathrm{FCF}_{jt} = \sum_i \theta^I_{ijt} Y_{ijt},$$

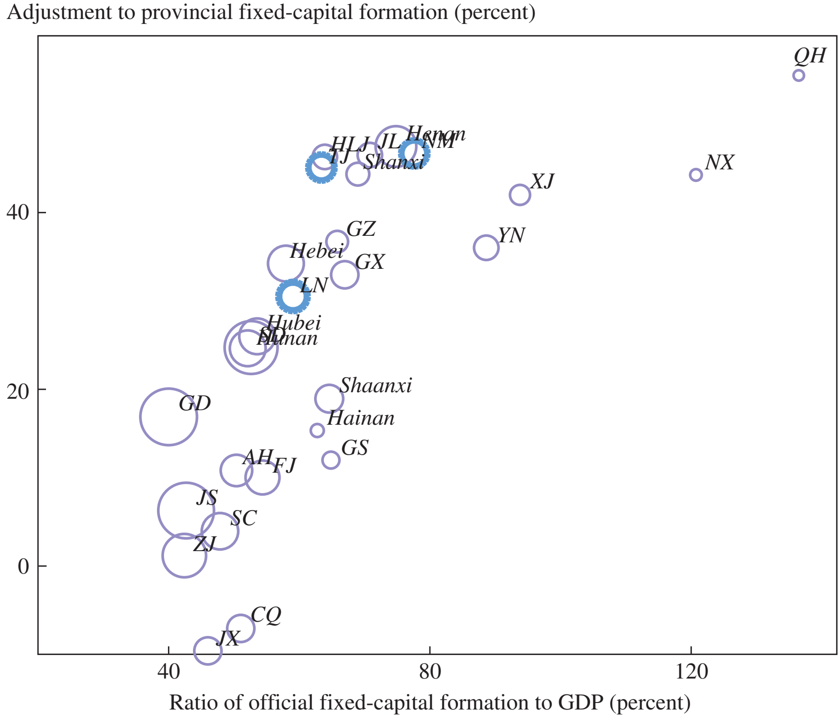
where j is the province index, and θ^I_{ijt} denotes the proportion of Y_{ijt}, sector i's value added in province j, that is converted to fixed capital in province j. Instead of using regional IO tables for θ^I_{ijt}, we assume $\theta^I_{ijt} = \theta^I_{it}$ and rely on the numbers in the national IO table. We then plot adjustment to provincial FCF against the official FCF GDP ratio in 2015 given in figure 12. We find that most provinces overreport FCF, and the extent of overreporting is increasing in the official investment rate. The overreporting of FCF is most severe in provinces such as Qinghai and Henan. The FCF GDP ratio was overstated by more than 50 percentage points in Qinghai. All three provinces discussed above, where local officials "confessed" to manipulating local statistics, are also associated with a severe overstatement of FCF. Their official FCF is about 30 to 50 percent higher than our estimates in 2015.[30]

Figure 13 shows a positive correlation between the extent of overreporting in provincial GDP and that overreporting in provincial FCF (the correlation is about 0.54). Although our estimated provincial GDP

29. We calculated these two numbers from the published tabulations in the *China Statistical Yearbook*.

30. We use the 2014 FCF data for LN because FCF in LN declined by about 30 percent in 2015. Without a big adjustment in GDP, LN's net exports jumped from –104 billion yuan in 2014 to 304 billion yuan in 2015. In other words, before its GDP adjustment in 2016, LN had scaled back its investment in 2015.

Figure 12. Adjustment to Provincial Fixed-Capital Formation in 2015[a]



Sources: Authors' calculations, based on data provided by the National Bureau of Statistics of China and published tabulations from the *Input-Output Tables of China* and the *China Taxation Yearbook*.

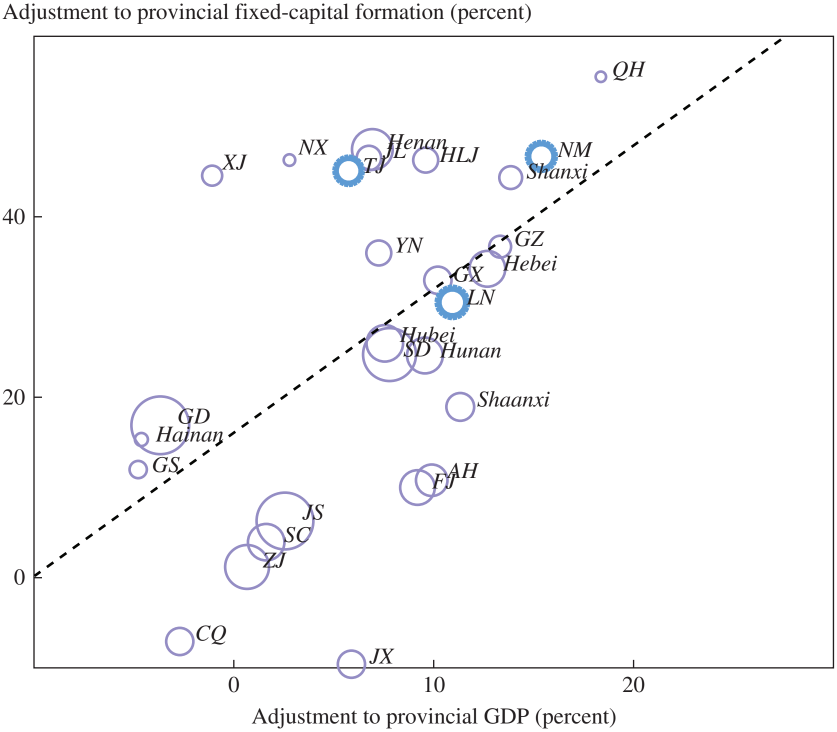
a. This is a scatterplot (FCF = fixed-capital formation) of (official provincial FCF – adjusted provincial FCF) / official provincial FCF (*y* axis) versus official provincial FCF / GDP (*x* axis).

and FCF are correlated by construction, there is no reason that the adjustments to provincial GDP and FCF should be correlated. If measurement errors in provincial GDP and FCF are large and independent, adjustments to the two variables would be uncorrelated. Figure 13 thus provides evidence that local governments overstate both GDP and FCF simultaneously.

II.C. Adjusting National Accounts with Statistical Models

A second approach is to explore the statistical relationship between GDP and a set of economic indicators outside China's national accounts. We first estimate a model using the provincial-level data before 2008, and we then use the estimated model and the indicators to predict provincial and national GDP after 2008. The success of the statistical approach

Figure 13. Adjustment to Provincial GDP and Fixed-Capital Formation in 2015[a]



Sources: Authors' calculations, based on data provided by the National Bureau of Statistics of China and published tabulations from the *Input-Output Tables of China* and the *China Taxation Yearbook.*

a. This is a scatterplot (FCF = fixed-capital formation) of (official provincial FCF – adjusted provincial FCF) / official provincial FCF (*y* axis) versus (official provincial GDP – adjusted provincial GDP) / official provincial GDP (*x* axis). The dashed line is a linear regression line.

depends on three conditions. First, the indicators are informative about local economies and are unlikely to be manipulated. Second, local GDP growth data before 2008 are more reliable than afterward. And third, the statistical model is flexible enough to capture the rich heterogeneity across Chinese provinces. We discuss these three conditions in order.

Our indicators include satellite night lights, national tax revenue, exports and imports, electricity consumption, railway cargo volume, and new bank loans.[31] National tax revenue is collected by local governments but is

31. Using bank loans (not new bank loans) delivers similar results.

directly paid to the central government. Cheating on national tax revenue would incur fiscal losses and, hence, is unlikely to happen. Exports and imports are from customs data, which are hard to manipulate due to the symmetry of the customs data from China's trading partners. Electricity consumption, railway cargo volume, and new bank loans are from the so-called Keqiang Index, which Li Keqiang, China's current premier, used to monitor local economic performance when he was the Communist Party secretary of Liaoning Province.

We understand that overreporting of local GDP started in the late 1990s. So local GDP growth data before 2008 cannot be entirely reliable. Yet we also understand that GDP overreporting has become more severe since 2008. What we identify from the next exercise is the difference in the degree of GDP overstatement between the period before 2008 and the post-2008 period. Consequently, when we rely on local GDP growth data before 2008, which is per se likely to be overstated, to estimate the subsequent growth, our adjustment needs to be a lower bound. The true GDP growth might be even lower than our estimates for the post-2008 period.[32]

In terms of the statistical model, we use the method developed by Su, Shi, and Phillips (2016) to control for hidden economic structural heterogeneities across regions. Consider this linear model:

$$y_{i,t} = \beta_i' X_{i,t} + v_i + \varepsilon_{i,t},$$

where y_{it} is log GDP of province i at year t, $X_{i,t}$ is a $p \times 1$ vector of logarithm of the indicators, β_i is a $p \times 1$ coefficient vector, v_i captures provincial fixed effects and $\varepsilon_{i,t}$ is the independent and identically distributed error term with mean zero. In the special case where $\beta_i = \beta$ the model reduces to the standard fixed effects regression. The more general model can capture heterogeneous economic structures across regions. Intuitively, β for the regions where the local economy relies heavily on resources might be very different from the others. Specifically, we assume β_i to be group-specific—that is, $\beta_i = \beta_k$ for all i in group k, where $i \in \{1, 2, \ldots, N\}$, $k \in \{1, 2, \ldots, K\}$, and $K \leq N$. Instead of grouping provinces by geographical or economic

32. Our approach differs from that of Fernald, Hsu, and Spiegel (2015) and of Clark, Pinkovskiy, and Sala-i-Martin (2017), who use data on exports to China from its trading partners and night lights as independent measures of China's economic activities. We instead train our statistical model by provincial industrial GDP data prior to 2008, when the overstatement of industrial GDP was much less evident compared with the post-2008 period. Also see Hu and Yao (2019), who use night-time lights data to estimate GDP in a number of countries.

characteristics, we implement the classifier least absolute shrinkage and selection operator (C-Lasso) method used by Su, Shi, and Phillips (2016). The method provides statistical inference for membership identification, which is totally data driven. We do not have to rely on prior knowledge about the number of groups or the number of provinces within each group. With the groups identified from C-Lasso, we can use the fixed-effects model to estimate the group-specific coefficients.

It is worth mentioning the rapid expansion of China's service sector. According to the national accounts data, service accounted for about 43 percent of GDP in 2007, and the share increased to about 52 percent in 2017. This is important because some of our indicators, such as electricity consumption and railway cargo volume, might be more relevant for industrial production than for service production. If $y_{i,t}$ includes service output, the ongoing structural transformation would imply time-varying β_i and, hence, invalidate our model. To address the concern, we use provincial industrial GDP as $y_{i,t}$ in the benchmark and then use provincial GDP as a robustness check. There are two reasons why we prefer provincial industrial GDP. First, the stationarity of β_i is more defensible for industrial GDP alone. Second, we have shown the evidence that GDP overstatement is larger in the industrial sector.

Our sample consists of annual observations from 30 Chinese provinces (excluding Tibet) between 2000 and 2017. GDP, electricity consumption, exports and imports, railway cargo volume, and new bank loans are all from the NBS;[33] national tax revenue is from the *China Taxation Yearbook*; and we use the Defense Meteorological Satellite Program–Operational Linescan System night-time lights data from the U.S. National Oceanic and Atmospheric Administration.[34] The time series are shorter for some variables. Satellite night-lights data end in 2013. National tax revenue data end in 2015 because the reform to replace the business tax with the value-added tax made national tax revenue not comparable before and after 2016.

Two remarks are in order. First, night-light data, electricity consumption, and railway freight are all in real terms. As a robustness check, we use

33. All the indicators were downloaded from the NBS website. Exports and imports (by place of destination or origin in China) were priced in dollars and were converted into yuan by annual averages of the exchange rate. A new bank loan is the annual difference of the outstanding bank loan in December.

34. The night-light data are not comparable before and after 2010 due to the satellite change. We use the average of the light growth in 2009 and 2011 to proxy the 2010 light growth for out-of-sample predictions.

GDP deflators to convert GDP, national tax revenue, exports and imports, and bank loans into real terms in the regressions (see also Clark, Pinkovskiy, and Sala-i-Martin 2017).[35] The estimated GDP will be converted back into nominal terms. The online technical appendix reports the results with price adjustments. The differences are small. Second, we can use more data in the earlier period to estimate the model, with the caveat that the estimated model might be less applicable to recent years due to structural changes. In the online appendix, we estimate the model with data from between 1995 (the year after implementation of the tax-sharing reform) and 2007. The main results are very similar.

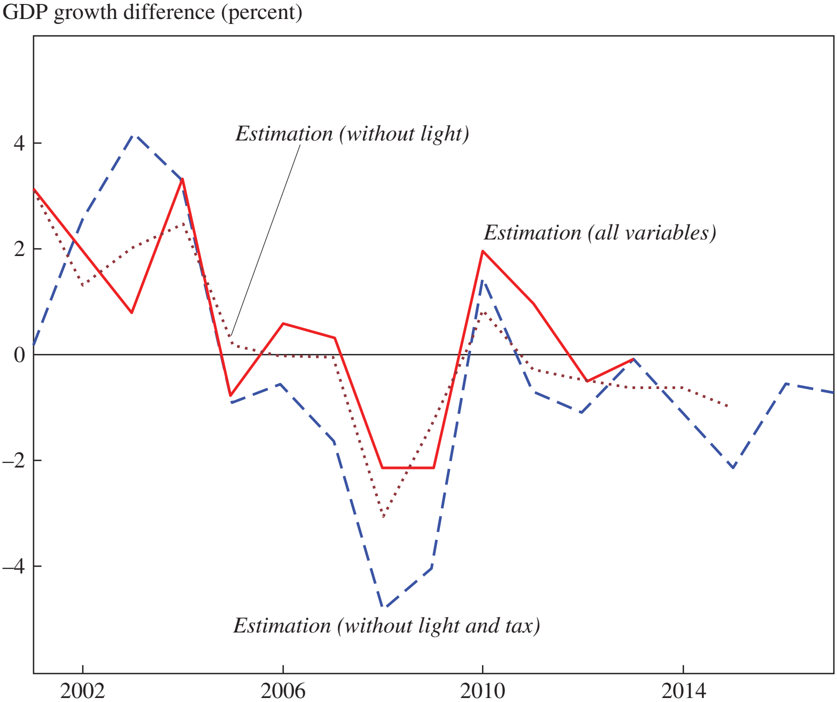
We first apply C-Lasso to the 2000–2007 data for model selection. K-fold cross validation, EBIC (Extended Bayesian Information Criterion), and data-driven penalty with heteroscedasticity (Belloni, Chernozhukov, and Hansen 2014; Belloni and others 2012, 2016) suggest keeping all the indicators except for new bank loans. Besides the statistical evidence, there is also an economic reason for us to drop bank loans. The "fiscal stimulus" launched by the Chinese government in late 2008 relaxed the borrowing constraint on local governments and led to a debt explosion afterward (Bai, Hsieh, and Song 2016). Much of the funds raised by local government financing vehicles are believed to finance infrastructure investment rather than production. This implies a structural change in the way that new bank loans contribute to GDP.

Our estimation is done in three steps. First, using the sample from before 2008, we run the C-Lasso estimation to classify provinces into different groups. Second, we estimate group-specific coefficients with post-Lasso ordinary least squares regressions. Finally, the estimated $\hat{\beta}_k$ and the same set of indictors are used to estimate provincial secondary industry value added throughout the whole sample period. Assuming that provincial agriculture, construction, and service GDP are reliable, we can estimate provincial GDP, which will be added together to obtain the aggregate GDP. Note that the estimated industrial value added after 2008 is out-of-sample prediction, while the estimation before 2008 is in-sample prediction.

When we use provincial industrial GDP, the C-Lasso procedure does not find statistical evidence for grouping, suggesting that the relationship between industrial GDP and these indicators is similar across provinces. As we show below, the result would be different if we replace provincial industrial GDP with provincial GDP. Because the satellite night-light

35. GDP deflators are inferred from the official real and nominal GDP growth.

Figure 14. The Gap between Estimated and Official Nominal GDP Growth, 2001–17[a]



Source: Authors' calculations.

a. This figure is based on the statistical model for provincial industrial GDP. It plots the difference between adjusted and official nominal GDP growth rates. "Estimation (all variables)" uses all variables. "Estimation (without light)" uses all variables except for night lights. "Estimation (without light and tax)" uses all variables except for night lights and national tax revenues. Adjusted GDP growth rates are in-sample estimates before 2009 and out-of-sample predictions after 2009.

data are not available after 2013, it can only be used for the out-of-sample prediction between 2008 and 2013. We rerun the C-Lasso and post-Lasso ordinary least squares regressions without night lights. The estimated model can make out-of-sample predictions for the post-2013 period.[36]

The out-of-sample predictions are shown in figure 14 and table 2.[37] Although the in-sample predictions are close to the official numbers, in

36. The tables with the regression coefficients are in the online appendix.

37. We aggregate provincial GDP growth by our estimated provincial GDP, which is based on the estimated provincial GDP growth and uses 2009 official provincial GDP as the benchmark.

recent years the out-of-sample predictions have been more volatile and lower than the official numbers. The estimated GDP growth is about 0.5 to 2.2 percentage points lower than the official GDP growth during 2014 and 2016 (see the dotted-and-dashed line in figure 14 and see the seventh and eighth columns of table 2).

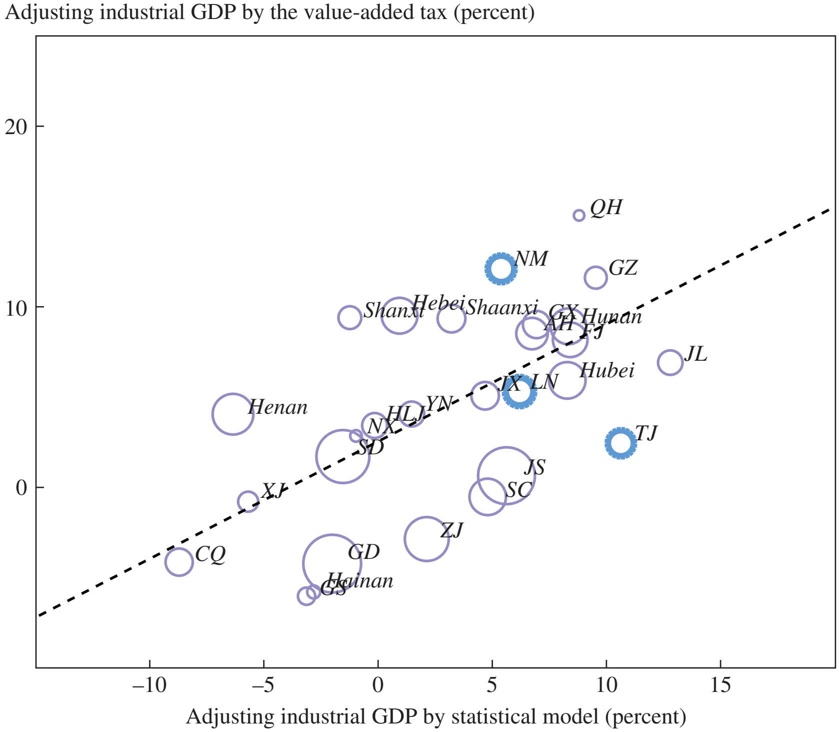
We note that although our two approaches are fundamentally different, they yield similar results in terms of the magnitude of overstatement of GDP. Table 2 shows that nominal GDP growth was overstated after 2010 and more so after 2013, and the magnitude of the overstatement after 2013 was about 1 to 2 percentage points.

One may wonder to what extent the tax revenue data used by the two approaches can explain their similar results on the recent overreporting of GDP. The first thing to notice is that tax revenue data are very different in the two approaches. National tax data include many taxes other than the value-added tax (for example, all consumption taxes and part of corporation income taxes), and only a fraction of value-added tax belongs to national taxes. Figure 15 plots the extent of GDP overstatement across provinces estimated by the first approach and the second approach with national tax revenue. Because the second approach only adjusts industrial GDP, we use the first approach that adjusts industrial GDP only to make the two approaches more comparable. The correlation is about 0.64. In other words, the different methods using different data sources deliver positively correlated estimates of provincial GDP overstatement.

We also run the regressions without national tax revenue. An advantage of dropping national tax revenue is to extend the estimation to the years after the completion of the reform to replace the business tax. The results are shown in figure 14 and in the last column of table 2. The overstatement of GDP growth after 2013 appears to be a robust finding, though its magnitude does depend on estimation method and variable selection.

We next replace provincial industrial GDP with provincial GDP for a robustness check. We drop both railway cargo volume and new bank loans, as suggested by Lasso. Given the huge disparity in GDP composition across provinces, not surprisingly, C-Lasso identifies two groups, with 16 provinces in group 1 and 14 provinces in group 2. See online appendix II for the detailed grouping results. Interestingly, Beijing, Shanghai, and Hainan, the three provinces with the highest service GDP share, are all in group 1. The fixed-effects regression results for each group are reported in the online appendix. Coefficients are indeed quite different across groups. We then run C-Lasso without night-light data, which also identifies two groups, with 11 and 19 provinces in groups 1 and 2.

Figure 15. Adjustment to Provincial GDP in 2015[a]



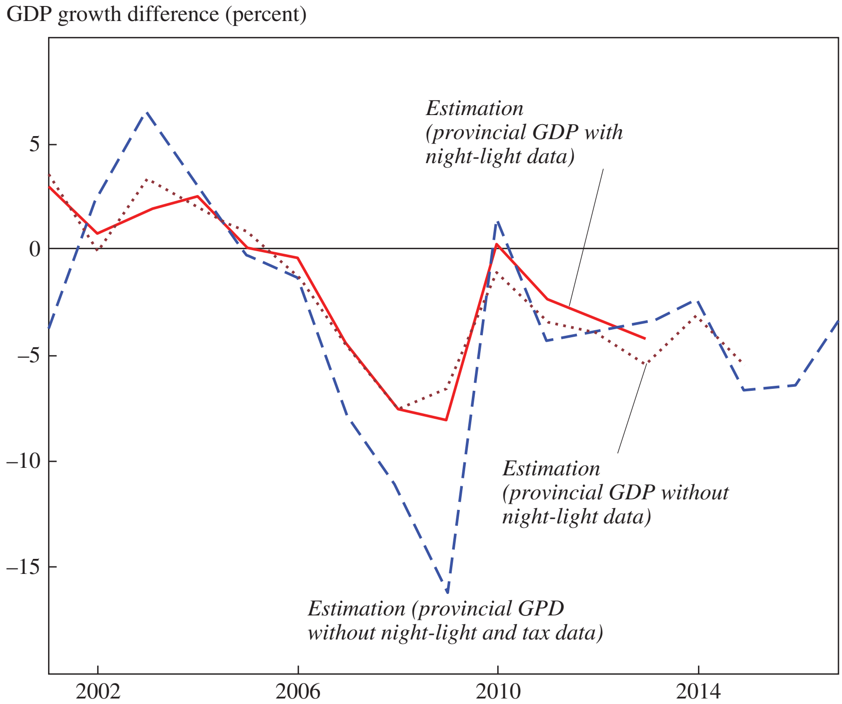
Source: Authors' calculations.

a. This figure is a scatterplot of (official provincial industrial GDP – estimated provincial industrial GDP by value-added tax) / official provincial GDP (y axis) versus (official provincial industrial GDP – estimated provincial industrial GDP by the statistical model with national tax revenue) / official provincial GDP (x axis). The dashed line is a linear regression line.

Online appendix II shows that 10 out of 11 provinces in group 1 are in group 1 identified by C-Lasso with night-light data. Again, Beijing, Shanghai, and Hainan are all in group 1.

Figure 16 compares the GDP growth rates from the official data, our estimates using provincial GDP with night-light data, provincial GDP without night-light data, and provincial GDP without night-light or tax data. Estimating provincial GDP directly implies a much bigger GDP overstatement. The difference between official GDP growth and our estimate was more than 5 percentage points in 2015. As discussed above, the caveat is the misspecification of the model that fails to capture how the rise of the service sector affects GDP growth.

Figure 16. Gap between Estimated and Official Nominal GDP Growth, 2001–17[a]



Source: Authors' calculations.

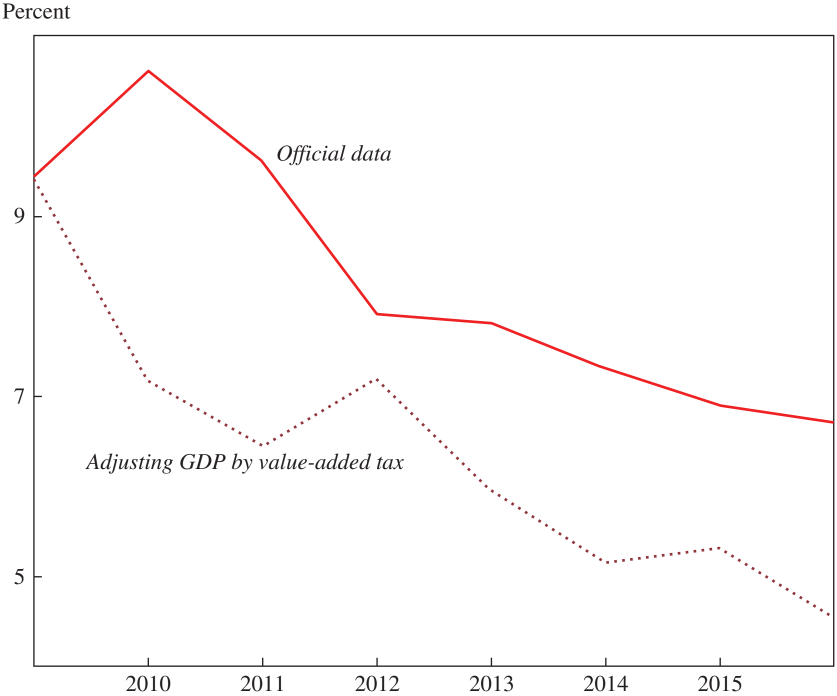
a. This figure is based on the statistical model for provincial GDP. See the note for figure 14. The only difference is that the estimation here is based on the statistical model for provincial GDP instead of industrial GDP in the province, as in figure 14.

III. Implications of Our Revised Estimates of China's National Accounts

To summarize the three main implications of our results: First, nominal GDP growth after 2010 and particularly after 2013 is lower than suggested by the official statistics. Second, the savings rate declined by about 11 percentage points between 2010 and 2016. The official statistics suggest the savings rate only declined by about 5 percentage points between these two years. Third, our statistics suggest that the investment rate fell about 8 percent of GDP between 2010 and 2016. Official statistics suggest that the investment rate fell 3 percent over this period.

We note that we do not have independent information on GDP deflators, so our statement is only about nominal GDP growth. The literature has

Figure 17. Real GDP Growth, 2009–16[a]



Sources: Authors' calculations; National Bureau of Statistics of China.

a. The adjusted real GDP growth is the difference between the adjusted nominal GDP growth rates (from the adjusted growth in the fourth column of table 2) and the growth rate of the official GDP deflator.

questioned the reliability of China's official price indexes, but we do not have independent information on the deflators.[38] Keeping in mind the caveats, we think it is useful to convert nominal output and input into real terms using the official GDP deflators and investment goods price index.

For real GDP growth, we calculate real GDP in the industrial, construction, and wholesale and retail trade sectors using our estimated nominal GDP (the first approach) and using the official GDP deflators for the three sectors. Adding adjusted real GDP in the three sectors to real GDP in the other sectors gives our adjusted real GDP, which is shown in figure 17. On average, the annual real GDP growth was overstated by about 2 percentage

38. See, for example, Brandt and Zhu (2010) and Nakamura, Steinsson, and Liu (2016).

points between 2010 and 2016. The official real GDP is about 13 percent above our estimate in 2016.

We now discuss the implications of our findings for capital returns, total factor productivity (TFP) growth, and the debt-to-GDP ratio. We begin with the return to capital. We use this equation to estimate returns to capital:

$$r(t) = i(t) - \hat{P}_Y(t) = \frac{\alpha(t)}{P_K(t)K(t)/P_Y(t)Y(t)} + \left[\hat{P}_K(t) - \hat{P}_Y(t)\right] - \delta(t),$$

where r denotes real returns to capital, i denotes nominal returns to capital, $\hat{P}_Y$ denotes the growth rate of output price, $\hat{P}_K$ denotes the growth rate of the capital goods price, α denotes the share of capital income in output, $P_K K/P_Y Y$ denotes the nominal capital-output ratio, and δ is the depreciation rate.

The results are plotted in figure 18.[39] The solid line uses the official data and replicates the earlier estimates made by Bai, Hsieh, and Yingyi Qian (2006) and the more recent ones by Bai and Qiong Zhang (2015). Recall that our adjustment of production GDP also lowers investment, which increases the ratio of output to capital. In either official or adjusted data, the dramatic decline in aggregate returns to capital in the post-2007 period turns out to be a robust phenomenon.

To estimate TFP, we assume this aggregate production function:

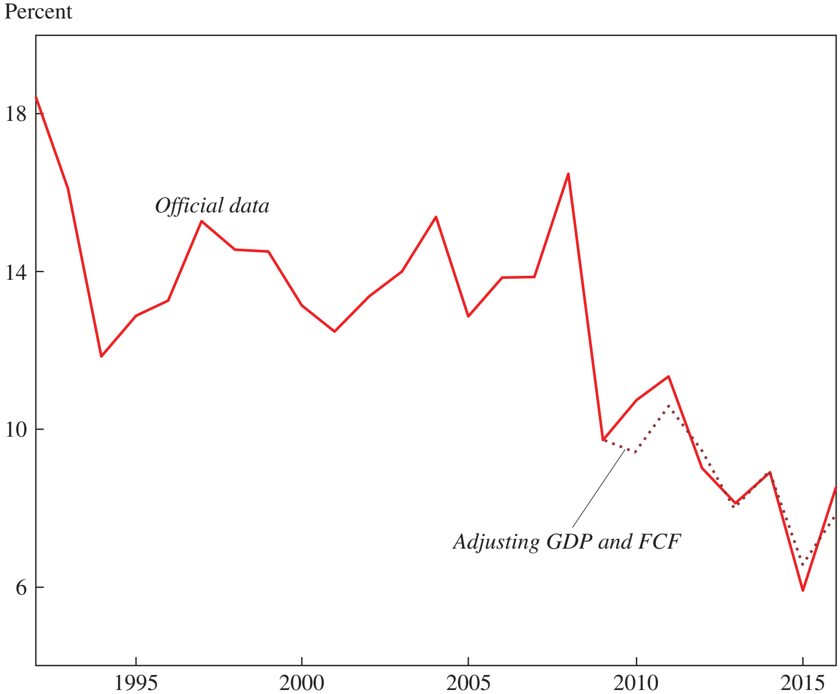
$$Y(t) = A(t)[K(t)]^{\alpha}[h(t)L(t)]^{1-\alpha},$$

where Y is real GDP, A is aggregate TFP, K is real capital, h is human capital per worker, and L is the number of workers.[40] The results are plotted in figure 19. The aggregate TFP growth rates given by our estimates appear to be more volatile than those given by official data. Yet it remains obvious that China's aggregate TFP growth slowed down substantially after 2007.

39. We discuss the details of the data used to estimate the return to capital in the online appendix.

40. We set $\alpha = 0.5$ (the results are similar if we use time-varying α calibrated in the online appendix for estimating returns to capital). We assume that $h = exp(s \cdot E)$, where E is the year of schooling and s is the return to schooling. The average year of schooling for workers in 2000 and 2005 is from the 2000 census and 2005 one-percent population survey data. We obtain the numbers between 2001 and 2004 by linear interpolation. For 2006 to 2016, we use the numbers from the labor force survey in the *China Population & Employment Statistics Yearbook*. For 1990 to 1999, we assume the annual growth of E to be its average growth from 2000 to 2005. We then use the 1 percent population survey data for 2005 to estimate returns to education by the Mincer earnings regression, which gives $s = 0.126$.

Figure 18. The Aggregate Return to Capital, 1994–2015[a]



Sources: Authors' calculations, based on the data provided by the National Bureau of Statistics of China and published tabulations from the *Statistical Yearbook of Chinese Investment in Fixed Assets, The Gross Domestic Product of China:* 1952–1995, and Hsueh and Li (1999).

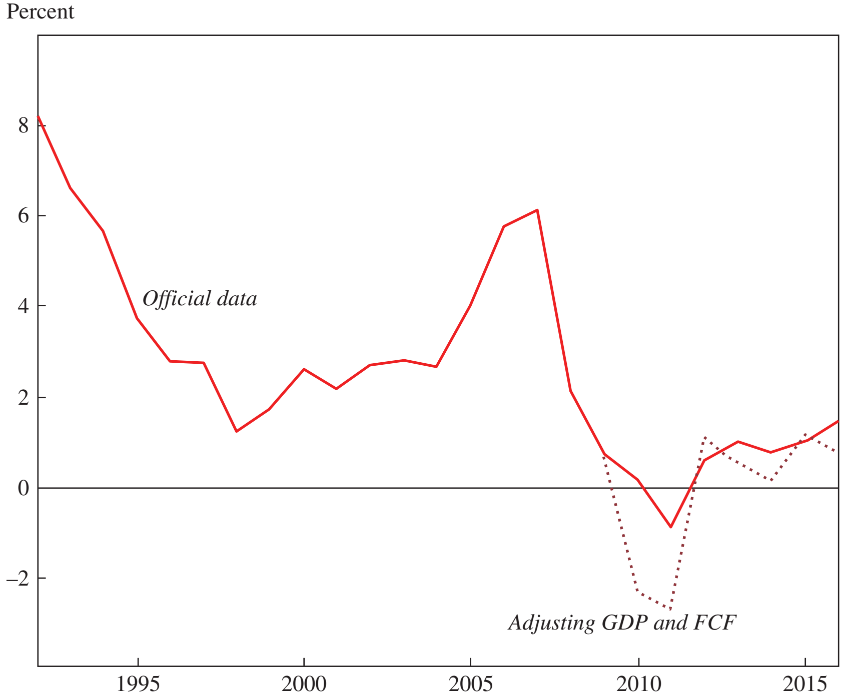
a. "Official data" indicate the return to capital computed from official GDP and fixed-capital formation (FCF). "Adjusting GDP and FCF" is the return to capital calculated from adjusted FCF and GDP (using the adjusted growth rate given in column 4 of table 2).

Finally, figure 20 shows the debt-to-GDP ratio with our revised estimate of nominal GDP. The estimation of debt follows that used by Song and Xiong (2018). The bottom line is that our revised numbers suggest that the debt-to-GDP ratio has increased by more than suggested by the official numbers. Our estimate of the debt-to-GDP ratio in 2016 is about 2.4—the official number is 2.1.

IV. Conclusion

A key institutional fact about China is that many administrative functions are controlled by powerful local governments. In another recent paper (Bai, Hsieh, and Song 2019), we argue that local governments have used this

Figure 19. Aggregate Growth of Total Factor Productivity, 1993–2015[a]



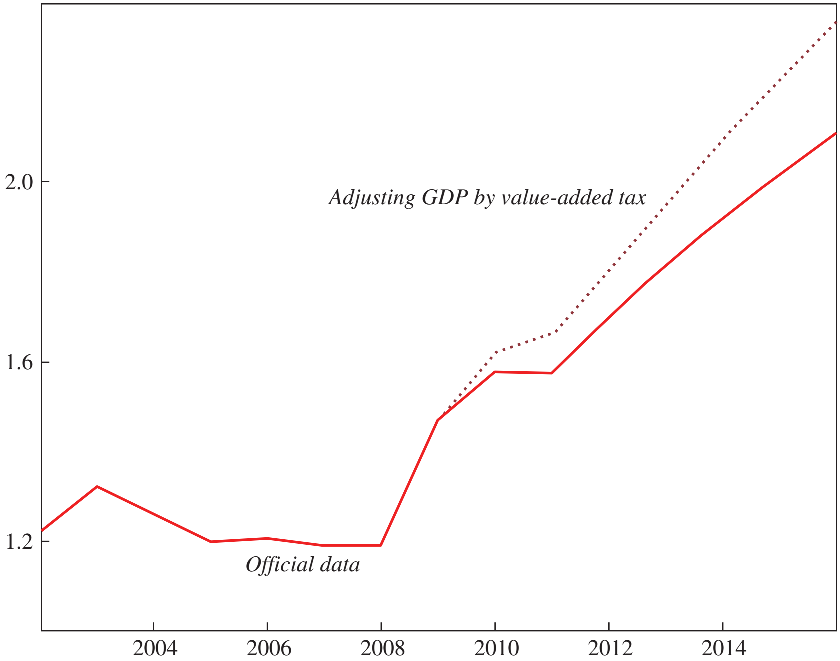
Sources: Authors' calculations, based on micro data from China's population censuses, data provided by the National Bureau of Statistics of China, and published tabulations from the *Statistical Yearbook of Chinese Investment in Fixed Assets* and the *China Population & Employment Statistics Yearbook.*

a. "Official data" indicate the return to capital computed from official GDP and fixed-capital formation (FCF). "Adjusting GDP and FCF" is the return to capital calculated from adjusted FCF and GDP (using the adjusted growth rate given in the fourth column of table 2).

power to support a large number of private businesses. The question in this current paper is what local governments choose to do with their power over local statistics.

We document that local governments have chosen to use this power to inflate local statistics on GDP, particularly by overstating industrial output and investment, particularly after the mid-2000s. As evidence, we show that the sum of local GDP has exceeded aggregate GDP since 2003. One possible explanation why they do this is the introduction of local economic performance in the evaluation of local officials by the Chinese Communist Party's Organization Department in the late 1990s. The official documentation of this policy change states that local officials will be evaluated based on "the speed, efficiency and potentials of economic

Figure 20. The Debt-to-GDP Ratio, 2002–16[a]



Sources: Song and Xiong (2018); National Bureau of Statistics of China.

a. The data on debt are from Song and Xiong (2018). "Official data" use official GDP in the denominator. "Adjusting GDP by value-added tax" uses adjusted GDP (using the adjusted growth given in the fourth column of table 2) in the denominator.

development, the growth of fiscal revenue, the improvement of people's living standards."[41] The revision intensified economic competition between local governments, and it seems likely that many local governments resorted to inflating local GDP numbers. Xiong (2018) provides a theoretical framework where competition between local governments results in the overstatement of both GDP and investment. And Changjiang Lyu and others (2018) present evidence that regional growth targets can be achieved by fabricating data.

The possibility that local governments misreport local GDP is well known, and the central government's National Bureau of Statistics adjusts the numbers reported by local governments. Before 2003, the NBS adjusted local GDP upward, but after 2003, it adjusted local GDP downward.

41. See Central Organization Department (1998).

However, our estimates suggest that the extent to which local governments exaggerate local GDP accelerated after 2008, but the magnitude of the adjustment by the NBS did not change in tandem. As a consequence, our best estimate is that the true growth rate of GDP is probably overstated by almost 2 percentage points from 2010 to 2016.

A final question is what tools and what incentives does the NBS have to report accurate statistics. We document that much of the underlying data behind the national accounts is out of the hands of the NBS. Furthermore, the question is what incentives does the NBS have to resist local officials who misreport data. Interestingly, although the NBS adjusts local statistics downward, it does not report the adjusted local statistics, perhaps out of a desire to not confront powerful local leaders. Given the NBS's weak position and the strong position of local leaders in the Chinese political system, it is not surprising that statistical data are potentially biased.

ACKNOWLEDGMENTS The authors thank David Dollar, Janice Eberly, Zhentao Shi, and Wei Xiong for helpful comments. Zheng Song acknowledges financial support from the Research Grant Council (Hong Kong) on "Re-Measuring China's Regional Investment," Project 14502718.

References

Bai, Chong-En, Chang-Tai Hsieh, and Yingyi Qian. 2006. "The Return to Capital in China." *Brookings Papers on Economic Activity*, Fall: 61–101.

Bai, Chong-En, Chang-Tai Hsieh, and Zheng Song. 2016. "The Long Shadow of China's Fiscal Expansion." *Brookings Papers on Economic Activity*, Fall: 129–81.

———. 2019. "Special Deals with Chinese Characteristics." *NBER Macroeconomics Annual*, forthcoming.

Bai, Chong-En, and Qiong Zhang. 2015. "The Return on Capital in China and Its Determinants." *China Economist* 10, no. 3: 20–37.

Bai, Chong-En, Xilu Chen, Zheng Song, and Xin Wang. 2019. "The Rise of China's Service Sector." Working paper.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Chris Hansen. 2012. "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain." *Econometrica* 80, no. 6: 2369–2429.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on Treatment Effects after Selection among High-Dimensional Controls." *Review of Economic Studies* 81, no. 2: 608–50.

Belloni, Alexandre, Victor Chernozhukov, Christian Hansen, and Damian Kozbur. 2016. "Inference in High-Dimensional Panel Models with an Application to Gun Control." *Journal of Business & Economic Statistics* 34, no. 4: 590–605.

Brandt, Loren, and Xiaodong Zhu. 2010. "Accounting for China's Growth." Working Paper 4764. Bonn: Institute for the Study of Labor. http://ftp.iza.org/dp4764.pdf.

Cai, Jing, and Ann Harrison. 2018. "Industrial Policy in China: Some Unintended Consequences?" *Industrial and Labor Relations Review*, Forthcoming.

Central Organization Department. 1998. "Provisional Regulations on Evaluating Party and Government Cadres." Document 6. Chinese Communist Party, Beijing.

Chen, Zhao, Xian Jiang, Zhikuo Liu, Juan Carlos Suárez Serrato, and Daniel Yi Xu. 2019. "Tax Policy and Lumpy Investment Behavior: Evidence from China VAT Reform." Working paper. https://economics.stanford.edu/sites/g/files/sbiybj9386/f/abstract_stanford.pdf.

Clark, Hunter, Maxim Pinkovskiy, and Xavier Sala-i-Martin. 2017. "China's GDP Growth May Be Understated." Working Paper 23323. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w23323.

Fernald, John G., Eric Hsu, and Mark M. Spiegel. 2015. "Is China Fudging Its Figures? Evidence from Trading Partner Data." Working Paper 2015-12. San Francisco: Federal Reserve Bank of San Francisco. https://www.frbsf.org/economic-research/files/wp2015-12.pdf.

Gao, Jie. 2016. "Bypass the Lying Mouths: How Does the CCP Tackle Information Distortion at Local Levels?" *China Quarterly* 228: 950–69.

Holz, Carsten A. 2013. "Chinese Statistics: Output Data." Working Paper 473. Stanford, Calif.: Stanford Center for International Development. https://globalpoverty.stanford.edu/sites/default/files/publications/473wp.pdf.

———. 2014. "The Quality of China's GDP Statistics." *China Economic Review* 30: 309–38.

———. 2018. "China's Investment Rate: Characteristics and Implications." Working paper. http://carstenholz.people.ust.hk/CarstenHolz-ChinaInvestmentRate-13May2018.pdf.

Hsueh, Tien-Tung, and Qiang Li. 1999. "China's National Income, 1952–1995." Boulder, Colo.: Westview Press.

Hu, Yinghao, and Jiaoxong Yao. 2019. "Illuminating Economic Growth." IMF Working Paper 19/77. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2019/04/09/Illuminating-Economic-Growth-46670.

Lardy, Nicholas R. 2014. *Markets over Mao: The Rise of Private Business in China*. Washington: Peterson Institute for International Economics.

Lei, Yu-Hsiang. 2017. "Quid Pro Quo? Government-Firm Relationships in China." Working paper. https://drive.google.com/file/d/181FwmoICvLQyWO5HOJhxu8_HcMLrLnHn/view.

Liu, Fang, Jun Zhang, and Tian Zhu. 2016. "How Much Can We Trust China's Investment Statistics?" *Journal of Chinese Economic and Business Studies* 14, no. 3: 215–28.

Lyu, Changjiang, Kemin Wang, Frank Zhang, and Xin Zhang. 2018. "GDP Management to Meet or Beat Growth Targets." *Journal of Accounting and Economics* 66, no. 1: 318–38.

Ma, Ben, Guojun Song, Lei Zhang, and David A. Sonnenfeld. 2014. "Explaining Sectoral Discrepancies between National and Provincial Statistics in China." *China Economic Review* 30: 353–69.

Ministry of Finance and the State Administration of Taxation. 2008a. "Detailed Rules for the Implementation of the Interim Regulations of the Value Added Tax." Document 50. Ministry of Finance and State Administration of Taxation, Beijing.

———. 2008b. "Notice of the Ministry of Finance and the State Administration of Taxation on Several Issues Concerning the National Implementation of Value Added Tax Reform." Document 170. Ministry of Finance and State Administration of Taxation, Beijing.

———. 2011. "Notice of the Ministry of Finance and the State Administration of Taxation on the Pilot Work of Levying the Value-Added Tax in Lieu of the Business Tax in the Transportation Industry and Some Modern Service Industries in Shanghai." Document 111. Ministry of Finance and State Administration of Taxation, Beijing.

———. 2013. "Notice of the Ministry of Finance and the State Administration of Taxation on Temporarily Exempting Certain Small- and Micro-Sized Enterprises from the Value-Added Tax and the Business Tax." Document 52. Ministry of Finance and State Administration of Taxation, Beijing.

———. 2014. "Announcement of the State Administration of Taxation on Relevant Issues Concerning Exempting Small and Micro-Sized Enterprises from the Value-Added Tax and the Business Tax." Document 57. Ministry of Finance and State Administration of Taxation, Beijing.

Ministry of Natural Resources. Various years. *China Land and Resources Statistical Yearbook*. Beijing: Ministry of Natural Resources.

Nakamura, Emi, Jón Steinsson, and Miao Liu. 2016. "Are Chinese Growth and Inflation Too Smooth? Evidence from Engel Curves." *American Economic Journal: Macroeconomics* 8, no. 3: 113–44.

National Bureau of Statistics. 2006. *Input–Output Tables of China 2002*. Beijing: China Statistics Press.

———. 2009. *Input–Output Tables of China 2007*. Beijing: China Statistics Press.

———. 2015. *Input–Output Tables of China 2012*. Beijing: China Statistics Press.

———. Various years. *China Population & Employment Statistics Yearbook*. Beijing: China Statistics Press.

———. Various years. *Statistical Yearbook of Chinese Investments in Fixed Assets*. Beijing: China Statistics Press.

Orlik, Tom. 2014. "Reform at China's National Bureau of Statistics under Ma Jiantang, 2008–2013." *China Economic Review* 30: 304–8.

Song, Zheng, and Wei Xiong. 2018. "Risks in China's Financial System." *Annual Review of Financial Economics* 10: 261–86.

Su, Liangjun, Zhentao Shi, and Peter C. B. Phillips. 2016. "Identifying Latent Structures in Panel Data." *Econometrica* 84, no. 6: 2215–64.

Xiong, Wei. 2018. "The Mandarin Model of Growth." Working Paper 25296. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25296.pdf.

Xu, Xianchun. 2004. "China's Gross Domestic Product Estimation." *China Economic Review* 15, no. 3: 302–22.

———. 2014. "Accurate Understanding of China's Income, Consumption and Investment." *Social Sciences in China* 35, no. 1: 21–43.

Zhang, Lei, Yuyu Chen, and Zongyan He. 2018. "The Effect of Investment Tax Incentives: Evidence from China's Value-Added Tax Reform." *International Tax and Public Finance* 25, no. 4: 913–45.

Comments and Discussion

COMMENT BY

DAVID DOLLAR This is an important paper. Every quarter, China's National Bureau of Statistics (NBS) reports data on GDP and its components, an event that is eagerly watched and that can move markets if the results are far from expectations. Buried toward the end of the Western press reports is usually a disclaimer. For example, the *New York Times* report on China's 2019 second-quarter growth stated that "economists widely doubt the veracity of the overall Chinese growth figure, which shows far more stability than comparable numbers from the United States and elsewhere" (Bradsher 2019). This paper aims to understand the main source of errors and to provide estimates of nominal GDP based on two different approaches—estimates that can be compared with official data.

This paper by Wei Chen and colleagues starts by examining how China's GDP is compiled. The NBS has branches all over the country, but these offices are integrated into local governments and have incentives to work with them. It has been known for some years that provincial GDPs, prepared by these offices, sum to significantly more than the national GDP reported by the central NBS. Clearly, the central NBS does not take the provincial estimates at face value and marks them down. This paper adds several pieces of useful information that help explain this process. First, overreporting of provincial GDP did not exist before about 2004; and second, it is almost wholly accounted for by overreporting of industry value added on the production side and by gross fixed-capital formation on the expenditure side. The authors tell a plausible story of local government officials since the mid-2000s being evaluated primarily based on implementation of big infrastructure projects and GDP growth, creating incentives for overreporting industrial value added.

The paper tries to come up with improved estimates based on two different methodologies. The first takes advantage of data on collection of the value-added tax (VAT). China has long had a VAT on goods, set at 17 percent. This is collected by finance bureaus at the local level, but they have no incentive to overreport it, given that most of it must be turned over to the central authorities. Starting in about 2008, the paper documents a tendency for VAT collected from industry to grow less rapidly than reported industrial value added. The authors use the VAT growth as a proxy for industrial value-added growth. This is the key input into their first alternative estimate. With this change, nominal GDP growth is reduced for the 2010–16 period from an official estimate of 11.50 percent a year to 10.37 percent a year (see Chen and colleagues' table 2, first and second columns). A similar adjustment to construction and wholesale and retail trade reduces estimated growth further, to 9.65 percent per year (the table's fourth column). Other services were taxed differently and in ways that changed over time, but the authors make a plausible case that estimates of service sector GDP were unbiased.

A potential problem with this first approach to adjusting GDP is that there were some changes in VAT rates over this period. The authors note that actual collection of VAT revenue increased from 10.4 percent of industrial value added in 2001 to 12.9 percent in 2007. The authors attribute this to improved tax collection. But it also likely reflects the fact that in practice, the tax rate was not always 17 percent for goods and that there were significant exemptions. In particular, China has long had VAT exemptions or rebates for exports, as is allowed under the World Trade Organization's rules. These rebates are not relevant because the authors use tax data gross of rebates. But in quite a few of China's export-processing zones, exports produced with imported components are exempted from the VAT. Shenzhen is a good example of a special zone that had this tax treatment (Ministry of Commerce 1986). Furthermore, China's policy toward the exemptions has changed over time. In the mid-2000s, as the overall trade surplus became very large and trade friction with the United States accelerated, China eliminated fully or partially the exemptions for many key exports. When the global financial crisis hit in 2008, the government reinstituted the VAT exemptions as a way of stimulating exports. The irregular application of the VAT was a constant complaint of the United States in economic dialogues with China at this time.[1] The issue

1. See, for example, USTR (2010, esp. 42–46).

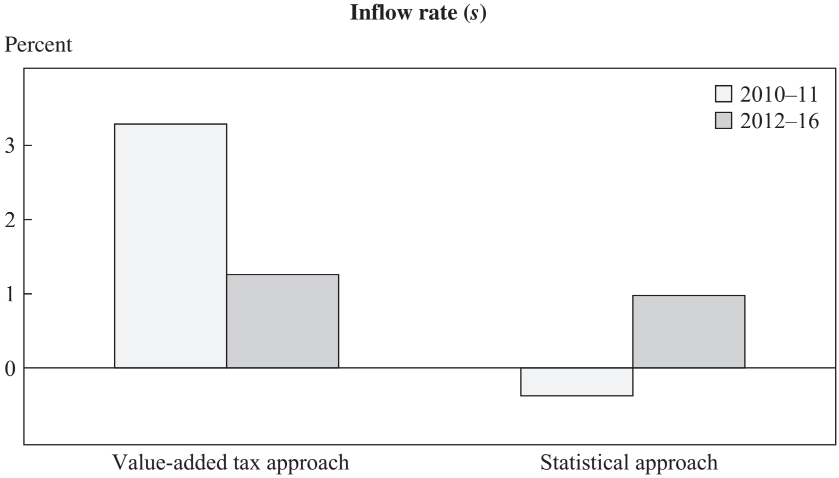
for the paper by Chen and colleagues is that there were changes in taxation policy and implementation right around the time of the global crisis.

Note in the top panel of Chen and colleagues' figure 4 that in the mid-2000s, VAT collected grew faster than industrial value added, but that starting with the global financial crisis, it grew more slowly. It is possible that some of the change resulted from changes in VAT rates for certain goods. In particular, as China started to recover from the global crisis, there were several years when industrial exports could increase without any increase in tax collection. In Chen and colleagues' figure 4, the largest discrepancy between the growth rate of VAT collection and of reported industrial value added is during 2009, 2010, and 2011, when China's stimulus was at its fullest.

A strength of the paper is that is has a second approach to estimating an alternative GDP series, which can then be compared with the results from the first approach. The second approach relies on regression analysis, specifically, a panel across provinces and over time. It assumes that industrial value-added data before 2008 were accurate, and it uses a set of explanatory variables whose data collection would plausibly not be subject to the same biases as the provincial reporting of industrial value added. These variables are national tax revenue collected from the province, satellite light data, electricity consumption, railway cargo, and trade. An equation is estimated for industrial value added with the data from 2000 to 2008. It is then used to predict provincial industrial value added until 2016. The aggregation of these provincial estimates is the estimate of national industrial value added. This estimate then replaces the official estimate of industrial value added to come up with a revised estimate of GDP. The light data are not available after 2013, and the tax data have the same potential problem already discussed. The authors include a reestimated GDP growth rate using this approach but dropping the light and tax data (Chen and colleagues' table 2, seventh column). For the period 2010–16, the reestimated growth rate averaged 10.88 percent a year, compared with the official figure of 11.50 percent a year and the reestimate from the first approach of 9.65 percent a year.

Thus, there is a pretty significant discrepancy between the results of the two approaches. The first approach suggests that GDP growth in China is overestimated by 1.8 percentage points a year; this is the number cited in the paper's abstract. The second approach finds a much smaller distortion, of 0.62 percentage point a year. It turns out that the difference in the estimates arises almost completely from the estimates for the crisis years, 2010 and 2011. For 2010–11, the VAT approach finds that official GDP

Figure 1. Overestimation of Official GDP Growth



Source: Author's calculations.

growth was overestimated by 3.3 percentage points a year; the statistical approach finds that official numbers *underestimated* actual growth by 0.4 percentage point a year (see my figure 1). Recall that these were years when the growth of VAT collection dropped sharply; but apparently the data from electricity consumption, rail cargo, and trade do not confirm such a sharp slowdown.

For the 2012–16 period, conversely, the two approaches provide very similar estimates. The VAT approach finds official GDP growth to be overestimated by 1.3 percentage point, compared with 1.0 according to the statistical approach. Hence, what is robust is that in recent years China's growth numbers have been overstated by about 1 percentage point.

A couple of empirical points from the paper change if the overestimation of GDP is taken from the statistical approach rather than the VAT tax approach. In particular, because the error in GDP on the expenditure side is in investment, the new estimates provide lower estimates of the investment rate and the savings rate than we see in official data. The paper reports that using the VAT tax-based estimates reduces the investment rate by 7 percentage points, to 35.5 percent of GDP in 2016. The savings rate is similarly reduced by about 7 percentage points of GDP, to 39.7 percent of GDP, because the balance of payments data are basically taken as accurate. If, alternatively, the revised estimates based on the statistical model are used, the changes to the investment and savings rates would be significantly smaller. A back-of-the-envelope calculation suggests that, using the estimates from the statistical model, the revision to the investment rate is only 2.4 percentage points of GDP. Given the uncertainty about each set of revisions, a reasonable approach would be to average the estimates. This approach would reduce the investment rate in 2016 by 4.7 percentage points of GDP. In light of this large amount of uncertainty, we can only conclude that the investment and savings rates are likely overestimated in official data, but it is difficult to say precisely by how much. Both the investment and savings rates have been trending down in the official data, so these revised estimates suggest that there has been even more rebalancing in the Chinese economy than previously thought—rebalancing in the sense of a shift from reliance on investment to a greater reliance on consumption on the expenditure side.

Another strong feature of the paper is that it is able to look at the issue of GDP overestimation province by province. Chen and colleagues' figure 15 shows the cumulative adjustment to provincial industrial value added, by 2015, according to both methodologies. A couple of things jump out from this figure. First, the big coastal provinces that produce much of China's GDP appear to have pretty good data; that is, there is no large distortion according to either methodology. This is true for Guangdong, Zhejiang, Jiangsu, and Shandong. Second, overall, there is a positive correlation between the results from the two approaches: the provinces found to have exaggerated GDP according to the first approach also have exaggerated GDP according to the second approach. Third, much of the overall distortion depends on 10 provinces in the upper-right quadrant of the figure, ones whose GDP is overestimated by at least 5 percent in 2015 according to both approaches. This group includes the rust belt of the Northeast (Heilongjiang, Jilin, and Liaoning) and the interior provinces (Qinghai, Inner Mongolia, Guizhou, Hunan, Hubei, Guangxi, and Anhui). A number of these provinces have had corruption scandals that have involved, among other things, falsified economic data. The point here is that if China wants to clean up this problem and improve its data, it knows where to start. The fact that the coastal provinces have generally reliable GDP data shows that China is capable of producing decent statistics.

In summary, this paper by Chen and colleagues is a useful examination of the likely errors in China's GDP statistics and GDP's major components on the production and expenditure sides. It makes a plausible case that local officials in some cases are able to get national acceptance of overestimation of provincial industrial value added and investment. A strength of the paper is that is uses two different approaches to reestimate industrial value added, and hence GDP. The two approaches yield significantly different estimates of the distortion in the GDP data, and the presentation emphasizes the larger distortion, when in fact the other approach is probably more defensible. What is robust in the two approaches is that the growth rate of nominal GDP has been overestimated by about 1 percentage point a year since 2010, that the savings and investment rates have come down more than is shown in official data, and that the errors are particularly large in 10 provinces of the Northeast and the interior.

REFERENCES FOR THE DOLLAR COMMENT

Bradsher, Keith. 2019, "China's Economic Growth Hits 27-Year Low as Trade War Stings." *New York Times*, July 14. https://www.nytimes.com/2019/07/14/business/china-economy-growth-gdp-trade-war.html.

Ministry of Commerce. 1986. "Supplementary Regulations for Tax Reduction and Exemption for Enterprises in Shenzhen Special Economic Zone, May 8." Ministry of Commerce, People's Republic of China, Beijing.

USTR (Office of U.S. Trade Representative). 2010. *Report to Congress on China's WTO Compliance.* Washington: U.S. Government Publishing Office.

COMMENT BY

WEI XIONG This paper by Wei Chen and his colleagues provides a fascinating study that systematically "corrects" China's GDP statistics. My discussion focuses on three aspects related to China's questionable GDP statistics: (1) the economic mechanism that leads to systematic misreporting; (2) the approach taken by Chen and colleagues to correct the misreporting; and (3) the potential consequences of such misreporting.

THE ECONOMIC MECHANISM OF OVERREPORTING Chen and colleagues provide compelling evidence of substantial inconsistencies between China's national and provincial GDP, mostly in the reported industrial output. Such inconsistencies reveal systematic overreporting in provincial statistics as well as the National Bureau of Statistics' (NBS) effort to correct the overreporting in national statistics. What causes the overreporting in provincial statistics? This overreporting problem is deeply rooted in the career incentives of provincial officials because (1) provincial bureaus of statistics report provincial economic statistics under strong influence from the provincial governments, and (2) some provincial statistics, such as provincial GDP and industrial output, are important measures in the central government's performance evaluations of provincial officials.

Even though the Chinese government has long abandoned central planning, it continues to play a central role in an increasingly market-driven economy. China has a complex government system, in which the central government works with regional governments at several levels: province, city, county, and township. Regional governments play key roles in China's economic development. Regional governments carry out over 70 percent of fiscal spending and are responsible for developing economic institutions and infrastructure at the regional level—for example, opening new markets and constructing roads, highways, and airports. Despite their autonomy in economic and fiscal issues, regional government leaders are appointed by the central government rather than being elected by the local electorate. As a key mechanism to incentivize regional leaders, the central government has established a tournament among officials across regions at the same level that uses economic performance to determine their career advancement. This system has greatly stimulated China's economic growth by giving local officials both fiscal budgets and career incentives to develop local economies. However, such powerful incentives may also lead to short-termist behaviors of local officials, such as overreporting local economic statistics, especially those that are most relevant for the performance evaluation.

My 2018 paper systematically examines such short-termist behaviors by developing a "Mandarin model" to account for the agency problems between China's central and local governments in affecting the Chinese economy. This model builds on the growth model of Robert Barro (1990) with a number of regions. In each region, the representative firm has a Cobb–Douglas production function with three factors: labor, capital, and local infrastructure. By creating more infrastructure in the region, the local government can boost the productivity of the local firm. Infrastructure investment represents the key channel for the local government to directly stimulate the local economy. The aforementioned tournament helps to mitigate the local government's tendency to underinvest in infrastructure relative to the social optimum. As more investment in infrastructure improves regional output, the tournament generates an implicit incentive for each region's governor to invest in infrastructure through the "signal-jamming mechanism" coined by Bengt Holström (1982), given that the central government is unable to fully determine whether regional output is due to the governor's ability or infrastructure investment. The powerful career incentives may also lead local governments to overreport regional output. This mechanism is similar in spirit to overreporting of earnings by executives of publicly listed firms (Stein 1989).

Specifically, the Mandarin model outlines a trade-off between the provincial government's GDP overreporting and the tax transfer to the central government. For the sake of argument, suppose that a provincial government collects tax revenue at a tax rate of τ of the province's GDP, Y_t, in year t, and needs to transfer a fraction of the tax revenue to the central government at a rate τ_c of the reported GDP Y'_t, with the transfer rate τ_c being lower than the gross rate τ. Then, the residual tax revenue for the provincial government is $T_t = \tau Y_t - \tau_c Y'_t$. It is clear that overreporting ($Y'_t > Y_t$) reduces the local fiscal budget, which in turn disciplines the overreporting. This trade-off leads provincial leaders with greater career incentives to overreport more. In addition, the Mandarin model implies that career incentives would also lead provincial leaders with strong career incentives to aggressively use leverage to boost their fiscal budgets at the expense of future debt burdens. Xiong (2018) provides a scatter plot of the ratio of provincial GDP overreporting to GDP, which is based on the estimation by Chen and colleagues and on the local debt-to-GDP ratio in 2015. The plot shows that these two types of short-termist behaviors are correlated across provinces. This curious association likely reflects the same mechanism driving these short-termist behaviors, as shown by the Mandarin model.

The dynamics of provincial GDP reporting are more nuanced than simply always overreporting. Commentators sometimes argue that fast-growing provinces like Guangdong have sufficient margins to meet their growth targets and may choose to underreport, rather than overreport, their GDP, because underreporting may help to reduce tax transfers to the central government and keep a strategic buffer for the future.

The inconsistencies between national and provincial GDP also reflect the fact that the central government is fully aware of the incentives of provincial governments to overreport their economic statistics and has made an effort to correct the overreporting in the national statistics. Interestingly, the NBS does not provide any breakdown on its assessments of overreporting by individual provinces, possibly because it does not want to publicly embarrass the provincial leaders, some of whom are already members of the Politburo and some of whom will eventually become national leaders. Furthermore, the correction by the NBS is constrained by its own data limitations, along with its own incentives. The NBS may not have a bias toward overreporting national statistics, yet it may be reluctant to report statistics that fall substantially short of the economic targets preannounced by the central government. In this sense, the NBS has also another set of incentives to manage the national statistics.

ESTIMATING OVERREPORTING Chen and colleagues use value-added tax revenue as the basis for correcting the overreporting in the GDP statistics. The premise of this approach is that local governments would not overreport value-added tax revenue because doing so would lead to greater shares of the actual tax revenues being transferred to the central government. This estimation approach is appealing and powerful and leads to a set of interesting and relevant corrections. Yet it is useful to note that although it is costly to overreport tax revenue, it may nevertheless happen in practice. In recent years, several regions—such as Liaoning, Tianjin, and Inner Mongolia—have officially acknowledged the previous overreporting of their GDP and tax revenues. These confessions typically happened after the previous regional leaders lost their prominence due to corruption allegations, and they helped the current leaders to obtain bigger fiscal subsidies or transfers from the central government. The confessed overreporting of tax revenue was substantial, even though it might have also been exaggerated due to the current leaders' incentives to plead for central government transfers.

CONSEQUENCES OF OVERREPORTING Economic statistics are an important source of information that helps policymakers, firms, and individuals adjust their policy, production, and investment decisions in response to time-varying economic conditions. It is difficult to directly estimate the potential distortions induced by the misreporting of economic statistics. China's Great Famine in 1959–61 was an extreme example of the deadly consequences of overreporting regional agricultural output. Xin Meng, Nancy Qian, and Pierre Yared (2015) provide forceful evidence that the severity of the famine was driven by overprocurements of grain during this period. Ziying Fan, Xiong, and Li-An Zhou (2016) point out a surprising observation: that the central government provided no famine relief during the first two years of the famine. To the contrary, China had exported a large quantity of grain to other countries, including sending food aid to African countries, in 1959 and 1960, after the famine had already spread across all of China. It is important to note that this devastating famine was also accompanied by widespread overreporting of grain output by regional leaders, due to their desire to support Mao Zedong's plan to quickly increase the country's agriculture production in support of the Great Leap Forward. Such overreporting might have led to excessive procurements of grain, which left insufficient grain to support local civilians. More profoundly, the overreporting might have also encouraged the central government's radical policies of moving more workers from agricultural production to industrial production just when grain output was rapidly falling.

In the modern era, there are many other sources of information for the central government and the general public to gauge the economic conditions in China. Thus, we do not expect this kind of fatal information breakdown to occur again. Nevertheless, the overreporting of economic statistics documented by this paper deserves close scrutiny.

REFERENCES FOR THE XIONG COMMENT

Barro, Robert J. 1990. "Government Spending in a Simple Model of Endogenous Growth." *Journal of Political Economy* 98, no. 5: S103–S125.

Fan, Ziying, Wei Xiong, and Li-An Zhou. 2016. "Information Distortion in Hierarchical Organizations: A Study of China's Great Famine." Working paper. https://wxiong.mycpanel.princeton.edu/papers/Famine.pdf.

Holmström, Bengt. 1982. "Managerial Incentive Problems: A Dynamic Perspective." In *Essays in Economics and Management in Honor of Lars Wahlbeck*. Helsinki: Swedish School of Economics.

Meng, Xin, Nancy Qian, and Pierre Yared. 2015. "The Institutional Causes of China's Great Famine, 1959–1961." *Review of Economic Studies* 82, no. 4: 1568–1611.

Stein, Jeremy C. 1989. "Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior." *Quarterly Journal of Economics* 104, no. 4: 655–69.

Xiong, Wei 2018. "The Mandarin Model of Growth." Working paper, Princeton University. http://wxiong.mycpanel.princeton.edu/papers/Mandarin.pdf.

GENERAL DISCUSSION Eswar Prasad observed that Chinese nominal GDP growth fluctuates while real GDP growth is smooth, which suggests both a problem with the deflators and any real GDP numbers derived using these deflators. Prasad noted a large change in the composition of Chinese GDP, of which the service sector accounts for more than 50 percent. This service sector growth has been largely in areas where the value-added tax (VAT) is much harder to collect. These two observations lead Prasad to believe that there might be an undercounting of GDP. He explained that a few years ago, the sampling frame for the service sector was rebased, and a few extra percentage points of GDP were found. He praised the authors for their analysis of the industrial and investment sectors, and noted that the Chinese household surveys give a reasonable sense of what is happening with household savings. In addition, he observed that there is already a reasonable measure of corporate and government savings, suggesting that it must be possible to reconcile these numbers with those generated by the authors to get the current account balance right. Finally, he mentioned his time working on China, where he became aware that even those working for the People's Bank of China were skeptical of the official GDP numbers. He recalled a former senior official who at one time became very concerned about the numbers of the Chinese National Bureau of Statistics (NBS) and came up with his own measure. He presented these new numbers—which differed significantly from the official numbers—to the State Council. Two weeks later, this same official received a letter from the NBS informing him that it was illegal for any other institution besides the NBS to collect data on GDP.

Jason Furman remarked that there have been two hypotheses about Chinese data. The first is the smoothing hypothesis, whereby China reports its GDP sooner than anyone else and does not revise its numbers. The second hypothesis is that China is in fact overstating its own growth. Furman wondered what forensics Chen and others had devised about the first hypothesis, because it could be confused with the second. Indeed, an unrealized permanent negative shock to the growth rate would result in the smoothing of GDP. Finally, Furman asked whether Chen and others had looked at the deflators issue, given that he has heard it plays into both hypotheses.

Tom Orlik described a time in 1998 when Premier Zhu Rongji noticed some falsification and embellishments in the Chinese statistical system. Orlik understood that after 1998, the Central Government and the NBS became more powerful relative to the local statistical bureaus, and he asked the authors whether his understanding was correct. If so, he wondered whether something had changed in the most recent period that had impeded or prevented the NBS from cross-checking the local area statistical agencies' numbers.

Orlik observed that in the last decade, China's credit growth has been rapid and that most of this growth has gone toward paying for investment spending. At face value, Orlik noted that this contradicts Chen and others' thesis that investment is being overreported in the national accounts. Orlik noticed that this contradiction is lessened by David Dollar's observation that credit growth is not going toward anything and that little investment is happening. However, Orlik questioned whether this contradiction is completely diminished. Finally, Orlik pointed out compositional issues with the VAT data and wondered whether it serves as a good proxy for growth.

Elaine Buckberg commented that the government officials in China and the provinces are powerfully responding to incentives for promotion to higher office. However, Chinese officials are also incentivized to attract transfers to the provinces to make sure they remain in office. Buckberg asked at what point the incentive to retain their position offsets the incentive to overstate GDP. Similarly, she agreed with Prasad's observation about the rapidly growing, largely private, service sector; and she pointed out that service sector underreporting will increasingly offset government officials' incentive to overstate GDP.

John Haltiwanger wondered whether it would be possible to examine the external validity of the paper, noting that in the United States, there are times when the survey data and administrative data do not match, particularly during periods of large cyclical variation. As such, Hatiwanger asked whether Chen and his colleagues would be able to replicate their analysis using another country. Haltiwanger then suggested replicating how the U.S. Bureau of Economic Analysis measures investment, instead of using survey data. Specifically, the bureau calculates investment flows by taking the nominal production of the capital goods industry, subtracting exports, and adding imports.

Stanley Fischer observed the small difference between the official numbers and what Chen and his colleagues find. Fischer commented that the difference seemed especially small, considering existing research that finds an underreporting of GDP of about 15 percent due to black market activity in Western economies.[1]

Jonathan Wright reflected on Yingyao Hu and Jiaxiong Yao's paper, which uses satellite measurements of electricity as a true output measure.[2] Hu and Yao also compared their generated numbers with those of the NBS, and observed a misreporting number roughly double that of Chen and colleagues. Wright observed that Hu and Yao do not use any Chinese data to generate this number.

Richard Cooper mentioned that the prime minister of China usually preannounces GDP growth to the National People's Congress in a speech. Cooper contemplated the pressure the NBS must feel to line up its numbers with this official government narrative. For example, Cooper described that two years ago, the prime minister announced that targeted GDP growth would be between 6.5 and 7 percent. Later, the head of the NBS clarified that its official 6.7 percent number rounded up to 7 percent. On a separate note, Cooper wondered about the validity of the official Chinese deflator numbers and advised the authors to address this issue in their paper.

1. Paulina Restrepo-Echavarria, "Macroeconomic Volatility: The Role of the Informal Economy," *European Economic Review* 70 (2014): 454–69, https://ideas.repec.org/a/eee/eecrev/v70y2014icp454-469.html.

2. Yingyao Hu and Jiaxion Yao, "Illuminating Economic Growth," working paper, 2019, http://www.econ2.jhu.edu/people/hu/paper_HUandYAO.pdf.

Jay Shambaugh described a time in about 2009 when the U.S. government was especially concerned that China was increasing its VAT rebates as a way to subsidize both exports and exporting industries. He wondered whether Chen and colleagues could look to see how much these VAT rebates on exports were increased, as an adjustment to their own adjustment.

Chang-Tai Hsieh clarified that their VAT number is the gross number before rebates. He noted that they chose to look at the industrial sector, in addition to wholesale and retail trade, because these are the two sectors to which the VAT tax applies. He noted that they decided to focus on these two industries because they were reasonably confident that there was relatively little evasion on the tax side and that any tax exemptions took the form of rebates on the export side. Although he wished they could do the same thing for the other service sectors, he noted that the VAT does not apply to these other sectors and that other data—such as corporate income tax data—were of poor quality and had more exemptions. As a result, he and his coauthors decided to use the official number for the other service sectors. He thanked Haltiwanger for his comment on a different way to measure investment data and noted that this method was implemented in their paper.

Hsieh explained that they used the official deflators and hoped that someone else would investigate whether the official deflator numbers were right. In response to whether the data on credit growth are consistent with their revised numbers for investment, he observed that an increase in credit does not necessarily translate into a one-to-one increase in investment, due to the substantial amount of substitution in sources of financing. In terms of the smoothing of GDP, he explained that nominal GDP becomes smoother after 2007. He noted that they did not feel confident that this smoothness really told them anything, due to the NBS's adjustments.

In response to comments made on whether the adjustments are big or small, Hsieh noted that it is hard to dispute that Chinese economic growth has not been high. He clarified that their paper only claims that GDP growth might be about 1 or 2 percent lower per year. Indeed, finding anything more—for example, that real GDP growth has been zero—would not be consistent with the obvious: that the Chinese economy has been growing. Hsieh clarified that he does not believe that the Chinese statistics are that different from the true number. Moreover, he explained that this discrepancy in the official numbers is similar to those in other countries. He referred to a paper he wrote on the Singaporean National Accounts, where he also found inconsistencies in the official data.[3] However, he noted that, although the difference found from the official number is small, he still believes it to be valuable information.

Finally, Hsieh observed the incredible amount of power held by local Communist Party secretaries. He noted that none of the central government ministers—such as the minister of finance—are members of the Politburo of the Communist Party. This implies that the party secretaries of Shanghai, Beijing, and Guangzhou, for example, are significantly more powerful than any finance minister or central bank governor. Of course, this power structure also means that the head of the NBS is vastly outranked by these local Communist Party leaders. He concluded that though this structure has contributed to the success of the Chinese economy, it also explains discrepancies found in the data and elsewhere.

3. Chang-Tai Hsieh, "What Explains the Industrial Revolution in East Asia? Evidence From the Factor Markets," *American Economic Review* 92, no. 3 (2002): 502–26.

RICHARD K. CRUMP
Federal Reserve Bank of New York

MARC GIANNONI
Federal Reserve Bank of Dallas

STEFANO EUSEPI
University of Texas at Austin

AYŞEGÜL ŞAHIN
University of Texas at Austin

A Unified Approach to Measuring u*

ABSTRACT This paper bridges the gap between two popular approaches to estimating the natural rate of unemployment, u_t^*. The first approach uses detailed labor market indicators, such as labor market flows, cross-sectional data on unemployment and vacancies, and various measures of demographic changes. The second approach, which comprises reduced-form models and dynamic stochastic general equilibrium models, relies on aggregate price and wage Phillips curve relationships. We combine the key features of these two approaches to estimate the natural rate of unemployment in the United States, using both data on labor market flows and a forward-looking Phillips curve linking inflation to current and expected deviations of unemployment from its unobserved natural rate. We estimate that the natural rate of unemployment was about 4.0 percent toward the end of 2018 and that the unemployment gap was roughly closed. Identification of a secular downward trend in the unemployment rate, driven solely by the inflow rate, facilitates the estimation of u_t^*. We identify the increase in labor force attachment of females, the decline in job destruction and reallocation intensity, and the dual aging of workers and firms as the main drivers of the secular downward trend in the inflow rate.

Conflict of Interest Disclosure: Richard Crump is vice president and function head of Capital Markets at the Federal Reserve Bank of New York; Stefano Eusepi is associate professor of economics at the University of Texas at Austin; Marc Giannoni is senior vice president and director of research for the Federal Reserve Bank of Dallas; and Ayşegül Şahin is the Richard J. Gonzalez Regents Chair in Economics at the University of Texas at Austin and an adviser to the Federal Reserve Bank of Dallas. Beyond these affiliations, the authors did not receive financial support from any 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 except the Federal Reserve Bank of New York and the Federal Reserve Bank of Dallas. The views expressed in this paper are those of the authors, and do not necessarily reflect those of the Federal Reserve Bank of New York, the University of Texas at Austin, or the Federal Reserve Bank of Dallas.

The unemployment rate in the United States peaked at 10.2 percent in October 2009. Since then, it has declined gradually, reaching below 4 percent for the first time in almost 20 years. A debate has arisen about how sustainable these low levels are and how monetary policy should respond. Starting with Milton Friedman (1968) and Edmund Phelps (1967, 1968), both academics and policymakers have endeavored to measure a sustainable level of unemployment and what implications deviations from this level have for price and wage inflation. This natural rate of unemployment, u_t^*, is broadly defined as the unemployment rate such that, controlling for supply shocks, inflation remains stable. The measure u_t^* is thought to vary over time with changes in the economy, such as demographic shifts, changes in the structure of the labor market, and technological advances.[1]

There are two popular approaches to estimating u_t^* in the literature. The first approach uses detailed labor market data, such as changes in demographics (Perry 1970; Summers 1986; Shimer 1998; Brauer 2007; Barnichon and Mesters 2018), labor market flows and job vacancies (Blanchard and Diamond 1989; Daly and others 2012), firms' recruiting intensity (Davis, Faberman, and Haltiwanger 2013) and skills mismatch (Şahin and others 2014). One potential limitation of this approach is the absence of information from inflation to infer u_t^*; moreover, these measures are not additive, because they cannot be considered as independent from each other and thus they are not conclusive as to the *level* of the natural rate of unemployment. Finally, there is the need for detailed data sets (for example, to build mismatch indexes) that are available only for the more recent period.

The second approach—which comprises reduced-form models (Staiger, Stock, and Watson 1997; Laubach 2001; Orphanides and Williams 2002) and dynamic stochastic general equilibrium (DSGE) models (Galí 2011; Gertler, Sala, and Trigari 2008; Galí, Smets, and Wouters 2012)—relies mainly on price and wage Phillips curve relationships, together with model-specific assumptions on aggregate demand. This approach, in

1. For example, in the face of a spike in unemployment during the Great Recession, the modest decline in inflation was, in part, attributed to increases in mismatch unemployment, decline in firms' recruiting intensity, the extension of unemployment benefits, and uncertainty in economic conditions. This time variation in u_t^* is reflected in the time series of forecasted longer-run unemployment observed in survey data and projections by the Federal Open Market Committee; see figure C.1 in the online supplemental appendix. The online appendixes for this and all other papers in this volume may be found at the *Brookings Papers* web page, www.brookings.edu/bpea, under "Past BPEA Editions."

contrast, makes little use of detailed labor market information, and has been subjected to two sets of criticism. First, the natural rate estimates obtained from these models tend to be surrounded by a considerable degree of uncertainty, hampering their use for policy decisions. Second, the relationship between "economic slack" and inflation has been called into question since the financial crisis of 2007–8, because the strong rise in unemployment did not lead to a sizable and persistent decline in inflation.

We combine the key features of these two approaches and estimate u_t^* using a forward-looking Phillips curve linking inflation to current and expected deviations of unemployment from its *unobserved* natural rate. This estimation relies on two key pieces of information. First, we propose a measure of the secular trend in the unemployment rate obtained from separation (unemployment inflow) rates and job-finding (unemployment outflow) rates. We exploit the rich cross-sectional variation in the flow rates of different demographic groups to obtain an estimate of the trends. Our analysis of unemployment flows identifies the downward trend in the inflow rate as the main driver of the secular unemployment trend. The identification of such trends aids the measurement of the unobserved natural rate of unemployment. Second, we use survey-based professional forecasts to measure the term structure of inflation expectations, that is, the forward-looking component of the Phillips curve. We find that it is vital to account for the behavior of expectations to reconcile the observed behavior of inflation and slack over time consistent with the research of Marco del Negro, Marc Giannoni, and Frank Schorfheide (2015) and of Carlos Carvalho and others (2017).

We estimate the natural rate of unemployment for the United States over the period 1960–2018. As of the third quarter of 2018, we estimate that u_t^* was about 4 percent; in particular, using only information from price inflation, we estimate that u_t^* stood at 4.0 percent, with a 68 percent confidence interval of 3.5 to 4.5 percent. When we add information from wage inflation, the estimate shifts down slightly, to 3.8 percent, with an associated confidence interval of 3.5 to 4.2 percent. We find that the unemployment gap was roughly closed by the end of 2018, as short-term inflation expectations approximately converged toward their long-run mean. More generally, we find that the natural rate of unemployment, estimated using both price and wage inflation, was steady, at just below 6 percent, in the 1960s; rose sharply in the 1970s, to over 8 percent; and then fell steadily, to below 5 percent, in 2000. During the 2000s up until the Great Recession, the natural rate of unemployment was range-bound. In the Great Recession, we document a rise in u_t^* of about 1 percentage point relative to its prerecession levels. We demonstrate that this estimate aligns well with estimated contributions to the unemployment rate attributed to mismatch unemployment and changes in recruiting intensity.

We trace the long-term decline in u_t^* over the last 40 years to a secular downward trend in the rate at which workers become unemployed—the inflow rate. The decline in the inflow rate reflects three important changes in the labor market: (1) the rise in participation and labor force attachment of females, which coincided with fewer labor force interruptions related to maternity and childbirth and culminating in the closing of the gender unemployment gap; (2) the shift of the labor force from younger workers, who frequently become unemployed, to older workers, who are less likely to become unemployed; (3) the aging of firms, as older firms tend to have reduced rates of job destruction (layoffs and firings). The second and third changes are connected, and we refer to them as the *dual aging* of the U.S. economy, which has resulted in less job destruction and unemployment incidence in the labor market, not only through a composition effect but also by reducing unemployment incidence (job destruction) for workers (firms) in all age groups. Dual aging stands out as an important driver of the lower trend rate of unemployment, especially in the last two decades. Together, these secular changes have reduced the overall flow rate into unemployment and, consequently, the unemployment rate itself.

The structure of the paper is as follows. Section I presents an overview of the paper and discusses its contributions relative to the extensive literature on the natural rate of unemployment. Section II estimates the secular trend in unemployment, using detailed information for unemployment inflows and outflows by demographic group. Section III introduces a simple forward-looking Phillips curve, discusses its theoretical underpinning, and details the estimation methodology. Section IV presents the time series for the natural rate of unemployment, u_t^*, for the sample 1960–2018. Section V provides a quantitative evaluation of three factors driving the trend decline in the unemployment inflow rate: the increase in female labor force attachment; the decline in job destruction and reallocation; and the dual aging of workers and firms in the economy. Section VI concludes.

I. Overview and Relation to the Literature

The object we seek to estimate is "the natural rate of unemployment," u_t^*, which is defined as the unemployment rate such that, controlling for supply shocks, inflation remains stable.[2] Although the relation between inflation and unemployment is a perennial topic in macroeconomics (Humphrey 1991), the concept of the natural rate is often attributed to Friedman (1968) and Phelps (1967, 1968), and the notation u_t^* can be traced back to Phelps. As originally suggested by Friedman, u_t^* is generally assumed to vary over time, possibly as a function of demographic shifts, changes in the structure of the labor market, or technological advances. Friedman, in his 1968 presidential address to the American Economic Association, wrote:

> To avoid misunderstanding, let me emphasize that by using the term "natural" rate of unemployment, I do not mean to suggest that it is immutable and unchangeable. On the contrary, many of the market characteristics that determine its level are man-made and policy-made. . . . Improvements in employment exchanges, in availability of information about job vacancies and labor supply, and so on, would tend to lower the natural rate of unemployment. (Friedman 1968, 9)

Friedman clearly pointed out changes in labor supply behavior and in the efficiency of the matching process in the labor market arising from better matching technology as shifters of the natural rate. However, despite this key insight, an ongoing assumption of the time was that the natural rate was about 4 percent, which caused policymakers to underestimate how tight the labor market was. Various influential papers in the inaugural volumes of *Brookings Papers on Economic Activity* in the early 1970s studied the rise in the *natural* rate of unemployment, such as those by Robert Hall (1970a, 1970b), Robert Gordon (1970a, 1970b), George Perry (1970, 1972), and Charles Schultze (1971). These papers emphasized the role of the changing demographic structure of the economy and the importance of labor market flows in assessing the natural rate in real time. Here, we expand on these enduring insights and estimate the secular trend in unemployment and integrate it into the New Keynesian Phillips curve.

2. The extensive literature on the natural rate of unemployment used long run, frictional, average, equilibrium, normal, steady state, lowest sustainable, Hodrick–Prescott trend, nonaccelerating inflation rate of unemployment (NAIRU) and the unemployment at full employment to refer to related, perhaps the same, object that we are trying to estimate. An insightful article by Richard Rogerson (1997), titled "Theory Ahead of Language in the Economics of Unemployment," discusses the confusion and uncertainty around the language used.

Our point of departure is a simple decomposition of the unemployment rate,

$$(1) \qquad u_t = \bar{u}_t + \underbrace{(u_t - u_t^*)}_{x_t} + \underbrace{(u_t^* - \bar{u}_t)}_{z_t},$$

where $\bar{u}_t$ is the *secular trend in unemployment* and u_t^* is the *natural rate*. The secular trend in unemployment, $\bar{u}_t$, captures the elements of the unemployment rate that are driven by slow-moving factors such as demographics and social change. The unemployment gap, x_t, measures the deviation of the observed unemployment rate from the natural rate and is the primary input to monetary policy considerations (for example, the goal of maximum employment). The natural rate of unemployment is defined as the sum of the secular trend component and a cyclical component z_t. Conceptually, we would expect the natural rate of unemployment to converge to $\bar{u}_t$ over time in the absence of shocks.

Although it is tempting to use traditional filtering techniques to eliminate the higher-frequency fluctuations in the unemployment rate, we instead rely on rich cross-sectional variation in unemployment flow rates by demographic groups to assess the extent of the secular trend in unemployment, $\bar{u}_t$. We do so for three main reasons: (1) the inherent asymmetry in the unemployment rate (Montgomery and others 1998; Hamilton 2005) makes it challenging to directly estimate its secular, slow moving trend; (2) the inflow/outflow dynamics of the unemployment rate—which is the source of the underlying asymmetry—by itself provides a better characterization of the evolution of the unemployment rate (Blanchard and Diamond 1990; Barnichon and Nekarda 2012; Şahin and Patterson 2012); and (3) extensive cross-sectional information on these flow rates enables us to better distinguish and analyze the underlying common and group-specific trends.

In estimating the secular trend in unemployment, we allow trends in unemployment inflows and outflows to vary by age and gender. This follows a long-standing body of literature dating back to George Perry's influential *Brookings Paper* in 1970, which recognized age and gender as the main demographic characteristics that need to be taken into account in assessing the natural rate of unemployment. In particular, Perry suggested an adjustment to account for the rising share of teenagers and females in the labor force that is often referred to as the *Perry-adjusted* unemployment rate. This adjustment—which assigns a lower weight to the unemployment rate of demographic groups with lower hours and wages—has been used in the literature in estimations of the Phillips curve such as those made by

Gordon (1982) and Lawrence Summers (1986); and it provides a basis for different measures of labor market underutilization, such as U-1 and U-6 (for definitions of these measures, see BLS 2018). Robert Shimer (1998) built on the research by Perry (1970), Gordon (1982), and Summers (1986), and provided a critical evaluation of the underlying assumption of applying demographic adjustments to the unemployment rate: demographic shifts in the labor market only affect the aggregate unemployment rate through the changing labor force shares without affecting group-specific unemployment rates. Shimer (1998) argued that this assumption is adequate with respect to changes in the age structure but is violated when there are changes in educational attainment. More recently, Regis Barnichon and Geert Mesters (2018) revisited the demographic adjustment of the unemployment rate and proposed a new demographic adjustment based on gross flows data. We build on the research of Barnichon and Mesters (2018) and examine the relationship between demographics and unemployment flows instead of focusing directly on the unemployment rate.

To connect inflation to the state of the labor market, we employ a forward-looking Phillips curve linking inflation to expected inflation and the unemployment gap. Following Friedman (1968) and Phelps (1967, 1968), and building on the rational expectations school of thought in the 1970s (Sargent 1971; Lucas 1972), it has become common to link the gap between the unemployment rate and a natural rate of unemployment to the inflation rate, through an expectations-augmented Phillips curve. According to this relationship, whenever the unemployment rate is equal to its natural rate, inflation and inflation expectations should settle to their long-run value in the absence of supply shocks. For this reason, the natural rate of unemployment is sometimes called the nonaccelerating inflation rate of unemployment (NAIRU).[3] Moreover, for given unemployment, inflation, and an assumption about inflation expectations, this relation allows for the estimation of u_t^*. We utilize survey-based expectations of inflation at different horizons to provide noisy signals of true inflation expectations and impose that the secular trend act as an anchor for the natural rate, although accommodating the possibility of persistent deviations.

A Phillips curve by itself is, however, not a panacea to estimate the natural rate of unemployment. Indeed, as many authors have emphasized,

3. Modigliani and Papademos (1975) defined the noninflationary rate of unemployment, which they referred to as NIRU, as a rate such that, as long as the unemployment rate is above it, inflation can be expected to decline, and they estimated it to be somewhat over 5.5 percent in 1975.

the estimates of the response of inflation to the unemployment gap in conventional backward-looking Phillips curves—that relate current inflation to a measure of economic slack and lags of inflation to proxy for inflation expectations—appear to have diminished substantially since the late 1980s (Hall 2011; Ball and Mazumder 2011). This raises two issues. First, an instability in key parameters of the Phillips curve renders the estimating of u_t^* more difficult. Second, relatively flat Phillips curves may result in uncertain estimates of u_t^*.

Several researchers (Ball and Mazumder 2011; Hall 2011; Blanchard 2016) have also questioned the Phillips curve relationships on the grounds that the dramatic increase in the unemployment rate and the collapse in economic activity recorded during the Great Recession should have implied a very large drop in the inflation rate, or even deflation, in contrast to the relatively modest decline in inflation registered in the aftermath of the Great Recession. However, recent research has shown that while the criticism of backward-looking Phillips curves is well justified, it does not apply to forward-looking Phillips curves linking inflation to the unemployment gap and expectations of *future* inflation.[4] For example, del Negro, Giannoni, and Schorfheide (2015) show that a relatively standard monetary DSGE model with forward-looking expectations and financial frictions can account remarkably well for the joint evolution of inflation and economic activity during and after the Great Recession. This is because forward-looking agents in the model understand that monetary policy will be more accommodative in the future the more activity contracts, thereby helping to anchor inflation expectations. Carvalho and others (2017) provide further evidence that it is possible to reconcile the observed behavior of inflation with the level of slack during the crisis and its aftermath. In particular, they show that inflation expectations have remained "anchored" over that period, which has contributed to promoting price stability.

We build on this insight and use a forward-looking Phillips curve linking inflation to the unemployment gap and expectations of future inflation, which is based on the model of Jordi Galí (2011). Importantly, the forward-looking nature of this Phillips curve implies that inflation depends not only on the contemporaneous unemployment gap but also on the entire path of

4. Early derivations of the New Keynesian Phillips curve that features expectations of future inflation were calculated by Calvo (1983), Rotemberg (1987), Roberts (1995), Fuhrer and Moore (1995), Yun (1996), Goodfriend and King (1997), Rotemberg and Woodford (1997), Galí and Gertler (1999), and Sbordone (2002).

expected future unemployment gaps. This tighter link, along with data on inflation expectations at various horizons, coupled with the secular trend in the unemployment rate, helps us identify u_t^*.

Finally, our examination of the determinants of the secular trend in the unemployment rate links to the recent but growing literature on the decline in labor market dynamism. Worker and job reallocation have declined substantially in the recent decades, as initially documented by Steven Davis and others (2006) and recently analyzed by Davis and John Haltiwanger (2014) and by Raven Molloy and others (2016). Our analysis links the decline in labor market fluidity to movements in the inflow rate and assesses its quantitative impact on the natural rate of unemployment.

II. The Secular Trend in Unemployment

We estimate u_t^* in two steps. In the first step, described in this section, we extract the slow-moving trends in the inflow and outflow rates using a linear state-space model and obtain the unemployment rate trend $\bar{u}_t$. In the second step, we combine this trend estimate—together with measures of price inflation, wage inflation, and inflation expectations—to infer the natural rate of unemployment from a New Keynesian Phillips curve.

It could be argued that, from a statistical standpoint, it is more efficient to jointly estimate the unemployment trend and the natural rate. Our choice reflects two main considerations. First, the current approach is simple to implement and is transparent. Single-step estimation would add significant complexity because it would require conducting inference with a nonlinear state-space model of a reasonably large dimension (see equation 3 below). Second, as we argue in section V, the evolution of the unemployment secular trend is driven by forces such as changes in labor supply behavior reflecting social change or slow-moving demographic changes. This is broadly consistent with the assumption that $\bar{u}_t$ evolves exogenously to the state of the business cycle or changes in monetary and fiscal regimes during our sample period.

Subsection II.A introduces and summarizes the flow origins of the unemployment rate. Subsection II.B characterizes overall flows into and out of unemployment, whereas subsection II.C focuses on these flows for specific demographic subgroups. Finally, in subsection II.D, we introduce a state-space model to estimate the slow-moving components of the inflows and outflows to unemployment that maps directly to the slow-moving component of the unemployment rate.

II.A. The Flow Dynamics of the Unemployment Rate

Our main premise is that the flow origins of unemployment rate movements help us better connect to the underlying drivers of unemployment fluctuations and trends. Therefore, we start with the evolution of the unemployment stock from month t to month $t + 1$:

$$dU/dt = s_t(L_t - U_t) - f_tU_t, \tag{2}$$

where L_t denotes the labor force, s_t is the inflow rate (separation rate) to unemployment, and f_t is the outflow rate (job-finding rate) from unemployment. Although s_t is generally referred to as the separation rate and f_t as the job-finding rate, we use the inflow/outflow terminology used by Michael Elsby, Ryan Michaels, and Gary Solon (2009) and by Elsby, Bart Hobijn, and Ayşegül Şahin (2010). This terminology creates a clear differentiation between s_t and f_t and employment-to-unemployment and unemployment-to-employment flow rates based on gross flows data computed using longitudinally matched monthly Current Population Survey (CPS) micro data.

The unemployment rate, u_t, is defined as the fraction of the labor force, L_t, that is unemployed: $u_t = U_t/L_t$. We follow Shimer (2005, 2012) and calculate the outflow probability F_t using the observation that

$$U_{t+1} - U_t = U^S_{t+1} - F_tU_t,$$

where U^S_{t+1} is the number of unemployed who report having been unemployed for less than five weeks. Solving for F_t,

$$F_t = 1 - \frac{U_{t+1} - U^S_{t+1}}{U_t},$$

which can be mapped into a Poisson outflow hazard rate

$$f_t = -\log(1 - F_t).$$

The idea behind this calculation is intuitive: individuals who reported being unemployed for less than five weeks were not in the unemployed pool in the previous month, and therefore subtracting them out from this month's unemployment pool leaves us with those unemployed persons who failed to exit unemployment between month t and month $t + 1$. Solving the differential equation 2 forward, as done by Shimer (2012), we can solve for the unemployment inflow rate s_t

$$U_{t+1} = \frac{(1 - e^{-[s_t + f_t]})s_t}{s_t + f_t} L_t + e^{-[s_t + f_t]} U_t.$$

Given the fast transitional dynamics of the unemployment rate in the United States—as noted by Shimer (2005); Elsby, Michaels, and Solon (2009); and others—the unemployment rate is closely approximated by its flow steady state value, given by

$$\frac{s_t}{s_t + f_t}. \qquad (3)$$

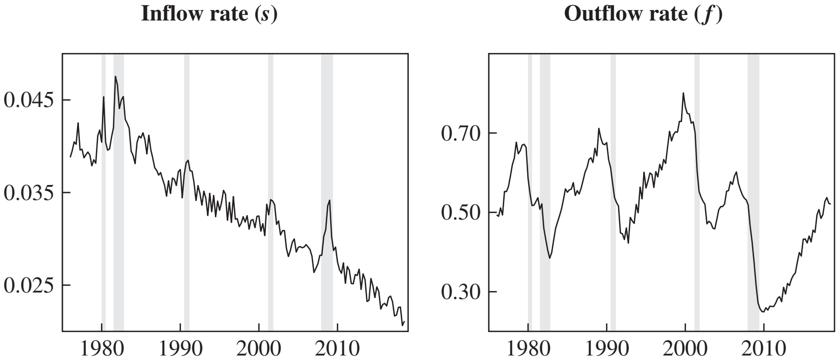
It is important to note that we focus on a two-state representation of unemployment, where we do not explicitly differentiate between the source of unemployment inflows and the destination of unemployment outflows, following Shimer (2005, 2012); Hall (2005); Elsby, Michaels, and Solon (2009); Elsby, Hobijn, and Şahin (2010); Davis and others (2010); and Şahin and others (2014). The inflow and outflow rates we use are estimated from CPS time series, rather than the longitudinally matched monthly CPS micro data. This abstraction simplifies the framework and better connects to the literature on unemployment dynamics. Although we maintain the two-state abstraction throughout section IV, we explicitly consider the role of the participation margin for females when we examine the drivers of the trends in unemployment flows in section V.

II.B. Unemployment Inflows and Outflows

The CPS provides us with monthly measures of the stock of unemployment, short-term unemployment, and the labor force. We calculate monthly unemployment inflow and outflow hazard rates using the methodology described above and plot quarterly averages of monthly s_t and f_t for the 1976:Q1–2018:Q4 period, as shown in figure 1.[5] Visual examination of inflow and outflow rates confirms the findings of the earlier literature regarding the cyclical properties of these flows. The inflow rate

5. Online supplemental appendix B provides details on the data sources used in this paper.

Figure 1. Inflow and Outflow Rates, 1976–2018[a]



Source: Current Population Survey.

a. This figure shows the unemployment inflow rate (left panel) and outflow rate (right panel) for the sample 1976:Q1–2018:Q4. Inflow and outflow rates are weighted averages of flow rates using counts of employed and not in the labor force from the six demographic categories described in subsection II.D.

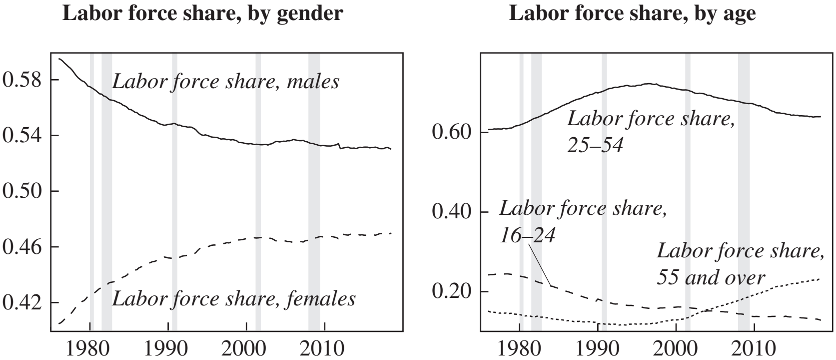
is characterized by sharp, short-lived spikes during recessions.[6] The outflow rate from unemployment is strongly procyclical, with persistent downswings during recessions.

Figure 1 also reveals that these two flows that shape the evolution of the unemployment rate over time exhibit differential long-run trends. The inflow rate has a striking downward trend, declining gradually to 0.02, half its level preceding the twin recessions of the early 1980s. In contrast, there is less evident trending behavior in the outflow rate.

Although it is tempting to use traditional filtering techniques to filter out the trends in the inflow and outflow rates, it is well known that the presence of a severe downturn—such as the Great Recession—at the end of the sample is likely to affect the estimate of the underlying trend. Instead, we rely on rich cross-sectional variation in the flow rates to assess the extent of the trends. In addition, cross-sectional information allows us to analyze the underlying drivers of the trends in the flow rates.

6. As emphasized by Shimer (2005, 2012) and Hall (2005), the response of the inflow rate was relatively muted during the mild recessions in 1990–91 and 2001. However, the inflow rate, without exception, exhibited sharp increases during severe recessions, including the most recent 2007–9 recession, as emphasized by Elsby, Hobijn, and Şahin (2010) and by many others.

Figure 2. Labor Force Shares, 1976–2018[a]



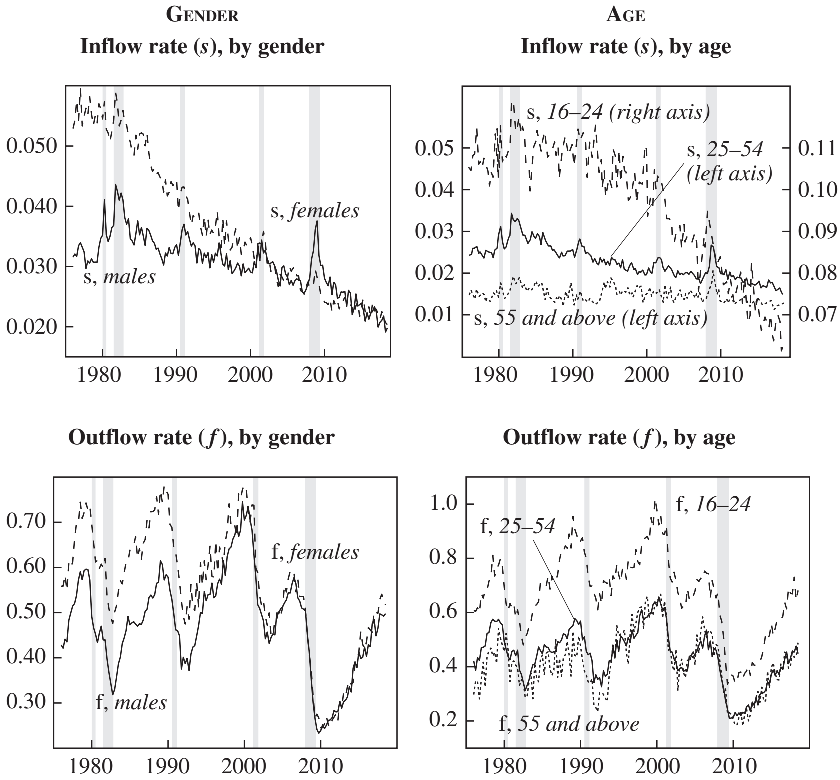
Source: Current Population Survey.
a. This figure shows the labor force shares by gender (left panel) and age (right panel) for the sample 1976:Q1–2018:Q4.

II.C. The Demographics of Unemployment Inflows and Outflows

We start with a visual examination of the flows by gender and age before we move on to our state-space setting to estimate the secular trends. Figure 2 summarizes the changes in the gender and age composition of the labor force from 1976 to 2018. With the rise in the female labor force participation rate, the labor force share of females increased to about 46 percent by 2000 and has stabilized. The age composition shifted from younger workers to prime age workers as the baby boom cohort entered the labor force and gradually aged, with the share of prime age workers peaking in the late 1990s. Since 2000, workers age 55 and older have constituted an increasing share of the labor force—and this has come about not just because of the aging of the population but also because of the differential trend in the participation rates of younger and older workers.

The left panels of figure 3 reveal the drastic convergence of males and females, both in terms of their unemployment inflow and outflow rates. A clear implication of this pattern is the disappearance of the gender unemployment gap, as discussed by Shimer (1998) and by Stefania Albanesi and Şahin (2018). Although the inflow rate has a downward trend for both male and female workers, the downtrend is more pronounced for female workers in the earlier part of the sample. The right panels of figure 3 show the importance of age composition. Workers younger than 25 years experienced an unemployment inflow hazard that was about five

Figure 3. Inflow and Outflow Rates by Gender and Age, 1976–2018[a]



Source: Current Population Survey.

a. This figure shows inflow rates (top row) and outflow rates (bottom row) by gender (left panels) and by age (right panels).

times that of those of prime age workers in the early 1980s. The recent decades show a partial convergence in their inflow rate as well as a decline in their shares.

II.D. Secular Trends in Unemployment Flows

In this subsection, we estimate the slow-moving trend in the inflow and outflow rates. We build on the research by Barnichon and Mesters (2018) and cast these trends in job market flows as latent processes in a linear state-space setting (Tasci 2014; Hornstein and Kudlyak 2019). Each flow is described by this set of equations:

$$s_t^i = \theta_i^s \varphi_t^s + \tau_t^{s,i} + \epsilon_t^{s,i} \text{ and} \tag{4}$$

$$f_t^i = \theta_i^f \varphi_t^f + \tau_t^{f,i} + \epsilon_t^{f,i}, \tag{5}$$

for $i = 1, \ldots 6$, our six demographic subgroups, and we normalize one element of θ_i^s and θ_i^f to 1. The individual flow rates are mapped into the aggregate rates s_t and f_t, using a choice of weights giving

$$s_t = \sum_{i=1}^{6} \omega_{it}^s s_t^i \quad \text{and} \quad f_t = \sum_{i=1}^{6} \omega_{it}^f f_t^i. \tag{6}$$

We use as weights for the inflow rate each group's share in the employment and out-of-the-labor-force pool because flows into unemployment originate from these two stocks.[7] For the outflow rate, we use each group's share in the unemployment pool. The trends in the inflow and outflow rate evolve according to

$$\tau_t^{s,i} = g_t^{s,i} + \tau_{t-1}^{s,i},\ g_t^{s,i} = g_{t-1}^{s,i} + \eta_t^{s,i} \text{ and} \tag{7}$$

$$\tau_t^{f,i} = \tau_{t-1}^{f,i} + \eta_t^{f,i}. \tag{8}$$

We assume that the slow-moving trend for the inflow rate, $\tau_t^{s,i}$, is characterized by an integration of order-two, $I(2)$, process, to accommodate the apparent secular trend in these flows. The trend for the outflow rate, $\tau_t^{f,i}$, is instead a random walk. The common component, $\varphi_t \equiv (\varphi_t^s, \varphi_t^f)'$, follows a second-order vector autoregressive process:

$$\varphi_t^f = \phi_{11}\varphi_{t-1}^f + \phi_{12}\varphi_{t-1}^s + \phi_{13}\varphi_{t-2}^f + \phi_{14}\varphi_{t-2}^s + \zeta_t^f \text{ and} \tag{9}$$

$$\varphi_t^s = \phi_{21}\varphi_{t-1}^f + \phi_{22}\varphi_{t-1}^s + \phi_{23}\varphi_{t-2}^f + \phi_{24}\varphi_{t-2}^s + \zeta_t^s. \tag{10}$$

The common cyclical component, φ_t, accommodates joint business cycle behavior between the inflow rate and the outflow rate for each group. This allows us to capture the specific lead-lag relationship around business cycle turning points (Fujita and Ramey 2009; Elsby, Hobijn, and Şahin 2013).

7. Recall from the text above that s is the solution to a nonlinear equation and that unlike f, it is not linear in group-specific weights. In unreported results, we verify that using employment shares and a more complex weighting method that corrects for time aggregation does not alter our main findings. However, small differences can arise between the overall series computed using aggregate data and the series constructed as a weighted average.

Finally, the flow-specific components follow a first-order autoregressive process

$$\epsilon_t^{s,i} = \psi_i^s \epsilon_{t-1}^{s,i} + \varepsilon_t^{s,i} \text{ and} \tag{11}$$

$$\epsilon_t^{f,i} = \psi_i^f \epsilon_{t-1}^{f,i} + \varepsilon_t^{f,i}, \tag{12}$$

representing idiosyncratic, possibly persistent, movements in the individual flow rates. The innovations $(\eta_t^{s,1}, \ldots, \eta_t^{s,6}, \eta_t^{f,1}, \ldots, \eta_t^{f,6}, \zeta_t^s, \zeta_t^f, \varepsilon_t^{s,1}, \ldots, \varepsilon_t^{s,6}, \varepsilon_t^{f,1}, \ldots, \varepsilon_t^{f,6})'$ are mutually independent, Gaussian random variables. The initial conditions $(\varphi_0, \varphi_{-1}, \epsilon_0^s, \epsilon_0^f)'$ are normally distributed with mean zero and unconditional variance implied by equations 9 through 12. Equations 4 through 6 represent the observation equations in the state-space model, and equations 7 through 12 are the transition equations. The model is estimated using Bayesian methods, utilizing the Gibbs sampler approach proposed by del Negro and others (2017).[8] (Also see Carter and Kohn 1994; Kim and Nelson 1999.) We estimate the model using quarterly data on labor market flows for the sample 1960:Q1–2018:Q3. We focus on six demographic subgroups: the interaction of three age groups—16–24 years, 25–54 years, and 55 years and over—with gender. Because individual flow rates are available only starting in 1976, we use aggregate flows for the remaining sample period, together with the weights, ω_{it}^s and ω_{it}^f, in order to estimate the unobserved trends for the entire sample.

The priors for the coefficients and variances of the vector autoregression with two lags, VAR(2), common components have the standard form

$$p\left(\text{vec}(\Phi)|\,\Sigma_\zeta\right) = \mathcal{N}\left(0, \Sigma_\zeta \otimes \Omega\right) \quad \text{and} \quad p\left(\Sigma_\zeta\right) = \mathcal{IW}\left(\kappa_\zeta, \Psi_\zeta\right), \tag{13}$$

where Φ is the autoregressive matrix corresponding to equations 9 and 10. The priors for the variance-covariance term of the innovations, Σ_ζ, is a fairly diffuse inverse Wishart with just enough degrees of freedom ($\kappa_\zeta = 4$) to have a well-defined prior mean for the innovations to the VAR. For simplicity of exposition, in this subsection only, we work with the flow rates multiplied by 100. The choices of priors in equation 13 imply standard deviations of 0.45 and 1.4 for the innovations to φ_t^s and φ_t^f, and also imply the mutual independence of these innovations.

8. We thank the authors for helpful discussions and for sharing their code.

The prior for Φ is a standard Minnesota prior with the hyperparameter for the overall tightness $\lambda = 2$ (which regulates the matrix Ω); this parameterization reflects a relatively loose prior. We implement similar priors for the two first-order autoregressive processes described by equations 11 and 12. The prior on the innovations corresponds to an inverse-gamma with shape and scale parameters implying a diffuse prior consistent with a well-defined mean.[9] The prior mean delivers a standard deviation of 0.45 for innovations to ε_t^s and 1 for innovations to ε_t^f. The prior on the autocorrelation coefficient is normally distributed with zero mean and variance determined using the same parametrization, with $\lambda = 2$ as in the VAR Minnesota prior. Also, in this case the prior is fairly diffuse. Finally, the priors on the innovations in the trend variables, $\tau_t^{s,i}$ and $\tau_t^{f,i}$, have inverse-gamma distributions. The parameters are chosen to guarantee a well-defined mean and deliver a standard deviation at the mean prior of 0.1 for the innovation to the trend in the outflow rate and of 0.01 to the innovation of the inflow rate.[10] The priors on the loadings θ^s and θ^f in equations 4 and 5 are defined as independent normal densities with mean 1 and standard deviation 0.5. Finally, we should note that in order to assess the role of the choice of priors, we have reestimated the model with uninformative priors and found broadly similar results.

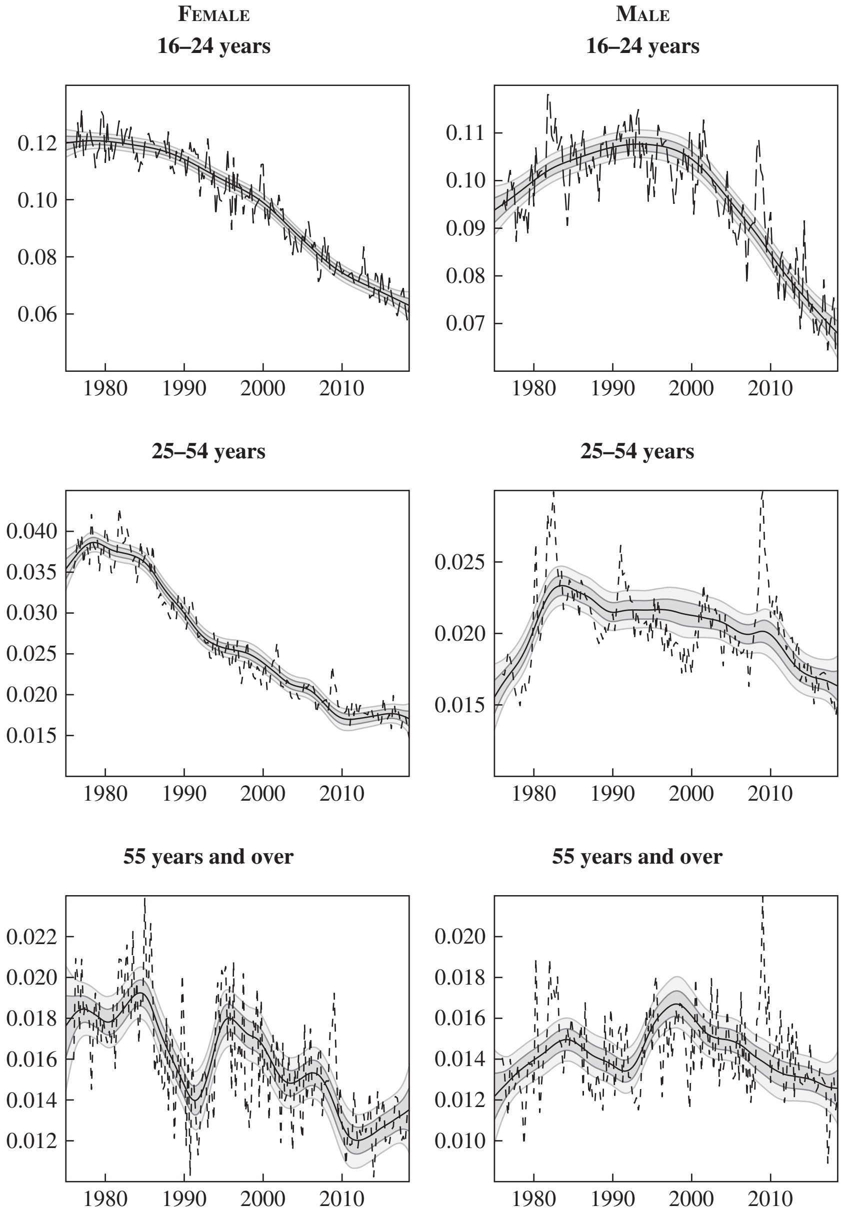
Figure 4 shows the six inflow rates and their estimated secular trend with associated confidence intervals, for the part of the sample where such flows are observable.[11] These trends differ substantially by gender and age group. Females age 16–24 and 25–54 show a pronounced downward trend starting in the 1980s, about halving their level from early in the sample. The trend for males age 16–24 displays a clear hump-shaped pattern, peaking in the first half of the 1990s and then falling by about 30 percent by the end of the sample. The remaining three groups, prime age males and females and males older than 55, demonstrate a milder secular trend. However, prime age males experience a notable decline in the inflow rate over the last decade or so. In contrast to the inflow rates, figure 5 shows that outflow rates have fairly stable trends, with the exception of females age 25–54. For this latter group, the outflow rate has fallen since the early 1990s. All other groups show little evidence of a secular trend over our sample.

9. The distribution has a shape parameter of 1.5 and scale parameters of 0.1 and 1 for ϵ_t^s and ϵ_t^f, respectively.

10. The gamma parameters correspond to a shape factor of 1.5 and a scale factor of 0.01 for the outflow rate and of $(0.01)^2$ for the inflow rate.

11. Individual flow rates going back to 1960 are estimated with a considerable degree of uncertainty. For this reason, we report below only the aggregate trend for the whole sample.

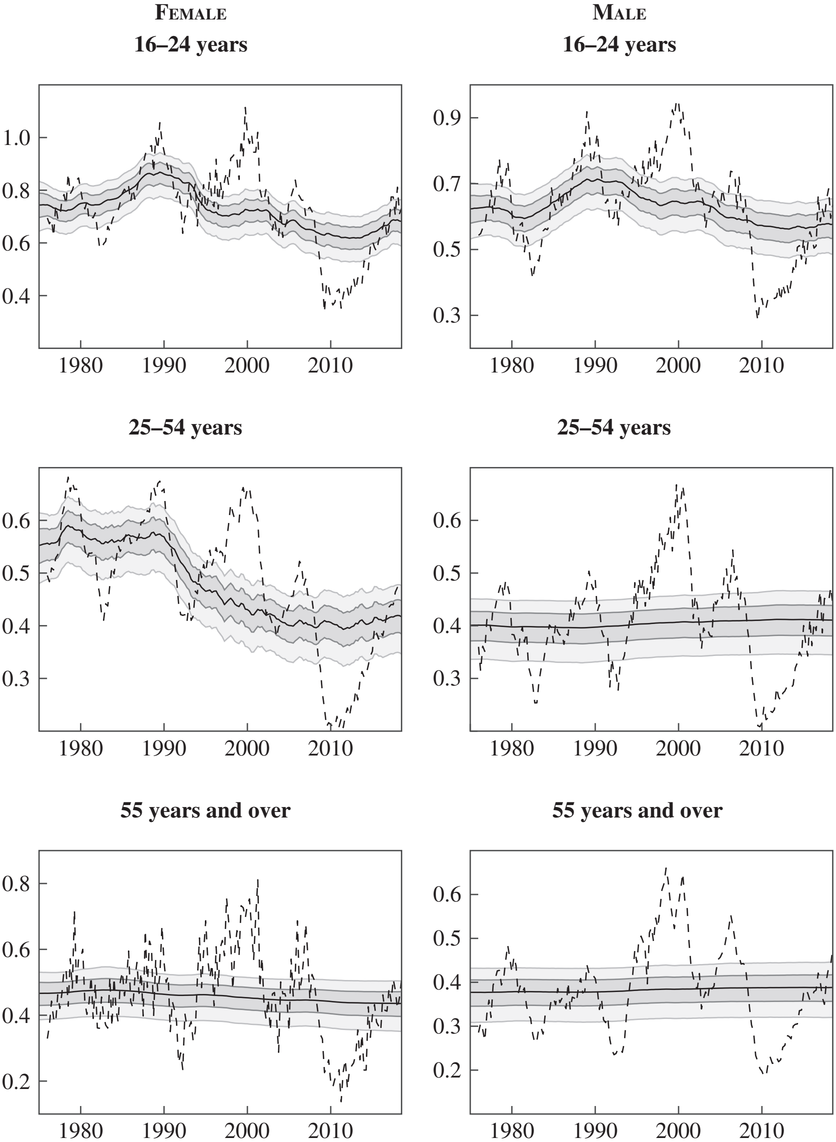
Figure 4. Inflow Rates by Gender and Age Subgroups, 1976–2018[a]



Sources: Current Population Survey; authors' calculations.

a. This figure shows observed inflow rates (dashed line) along with median estimates of the secular trend ($\tau_t^{s,i}$, solid line) for the inflow rate for each age and gender subgroup. Shading denotes 68 percent and 95 percent confidence intervals.

Figure 5. Outflow Rates by Gender and Age Subgroups, 1976–2018[a]



Sources: Current Population Survey; authors' calculations.

a. This figure shows observed outflow rates (dashed line) along with median estimates of the secular trend ($\tau_t^{f,i}$, solid line) for the outflow rate for each age and gender subgroup. Shading denotes 68 percent and 95 percent confidence intervals.

II.E. The Secular Trend in the Unemployment Rate

We map the individual secular trends for each subgroup using appropriate weights to obtain $\bar{s}_t$ and $\bar{f}_t$ as:

$$\bar{s}_t = \sum_{i=1}^{6} \omega_{i,t}^{s} \hat{\tau}_t^{s,i}, \qquad \bar{f}_t = \sum_{i=1}^{6} \omega_{i,t}^{f} \hat{\tau}_t^{f,i}. \tag{14}$$

Figure 6 shows the aggregate inflow rate, outflow rate, and unemployment rate along with their corresponding estimated secular trends—$\bar{s}_t$, $\bar{f}_t$, and $\bar{u}_t$—for the whole sample, 1960–2018. The secular trend in the inflow rate shows a decline of about 50 percent since the 1980s. In contrast, the secular trend in the outflow rate is generally stable, but has fallen since the 1990s, consistent with the evidence presented by Davis and others (2010). Finally, the secular trend in the unemployment rate, $\bar{u}_t$, can be constructed using $\bar{s}_t$ and $\bar{f}_t$ and the steady state approximation to the unemployment rate, via

$$\bar{u}_t = \frac{\bar{s}_t}{\bar{s}_t + \bar{f}_t}, \tag{15}$$

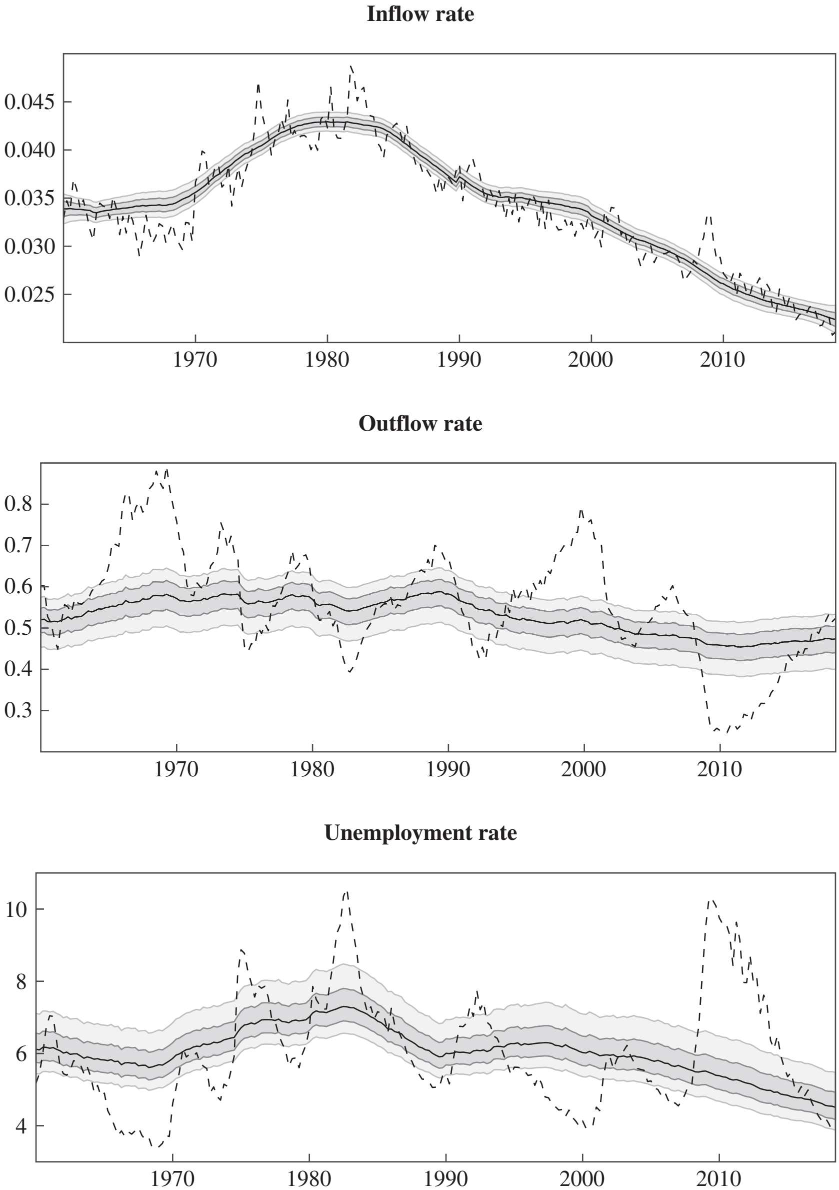
and is shown in the bottom panel of figure 6. The trend unemployment rate was about 6 percent in 1960 and increases to over 7 percent in 1983. Since then, it has displayed a clear downward trend, reaching about 4.5 percent by the end of the sample. Because the outflow rate shows little trending behavior, we observe from equation 15 that the overall downward trend is driven by the numerator, $\bar{s}_t$. The secular trend in the unemployment rate is estimated with a reasonably high degree of precision; for example, the 68 percent confidence interval at the end of the sample is comfortably less than 1 percentage point.

To illuminate interesting features of the trends in labor market flows over the last 60 years, we perform a number of counterfactual exercises using the model introduced above. Although this analysis is mostly descriptive, in section V we complement this analysis using more detailed micro data to analyze the economic drivers of these changes.

THE ROLE OF THE DECLINE IN THE OUTFLOW RATE As we observed in the previous subsection, $\bar{f}_t$ shows only a very modest decline in our sample. However, this decline has a nonnegligible role in the behavior of the trend in the unemployment rate. In the top panel of figure 7, we display two counterfactuals for $\bar{u}_t$:

(1) The measure $\bar{u}_t$ when $\bar{f}_t$ is set constant to its sample mean (the dotted line). We observe that, starting in 1990, we would have observed about half

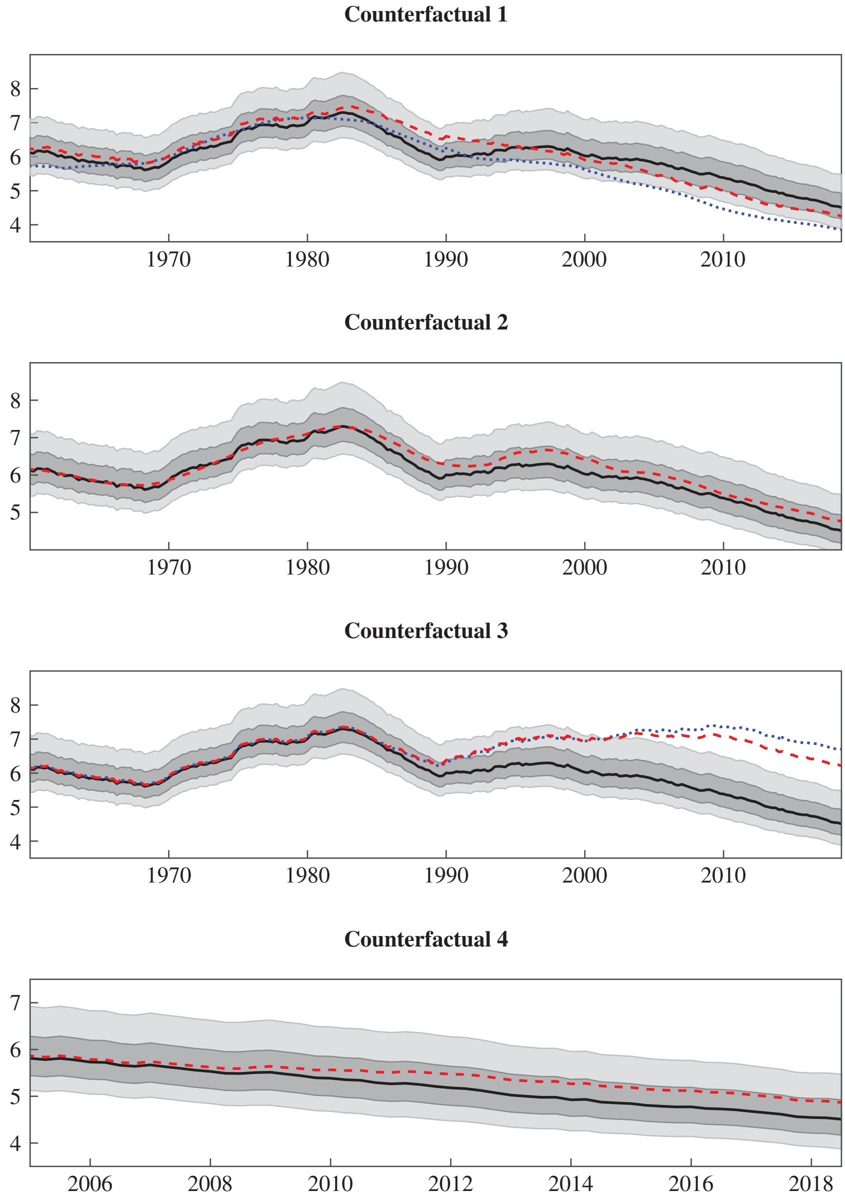
Figure 6. The Inflow Rate, Outflow Rate, and Unemployment Rate, Along with the Estimated Secular Trend, 1960–2018[a]



Sources: Current Population Survey; authors' calculations.

a. Actual rates denoted by dashed lines, median estimates of secular trend ($\bar{s}_t$, $\bar{f}_t$, and $\bar{u}_t$) denoted by the solid lines. Shading denotes 68 percent and 95 percent confidence intervals.

Figure 7. Counterfactual Exercises, Various Periods[a]



Sources: Current Population Survey; authors' calculations.

a. This figure shows the baseline estimate of $\bar{u}_t$ along with different counterfactual series (the dashed and dotted lines) based on the described scenarios. Shading denotes 68 percent and 95 percent confidence intervals.

a percentage point lower value of $\bar{u}_t$. This exercise implies that the entirety of the downward trend in $\bar{u}_t$ is driven by the inflow rate decline because it more than offsets the decline in the overall outflow rate.

(2) The measure $\bar{u}_t$ when $\bar{f}_t^i$ for $i = 1 \ldots, 6$ (the dashed line) are reestimated under the assumption that they are time invariant. We observe that this alternative tracks our baseline $\bar{u}_t$ closely, suggesting that $\bar{f}_t$ varies primarily through changes in the composition of the unemployed pool.

THE ROLE OF AGE AND GENDER COMPOSITION Section II summarized the sweeping demographic changes over the last 60 years that caused substantial shifts in age and gender composition. In this exercise, we capture the direct effect of the changing composition:

(1) The measure $\bar{u}_t$ when the weights $\omega_{i,t}^s$ and $\omega_{i,t}^f$ are fixed at their 1976 level (the dashed line). Between 1976 and the late 1990s, the change in the shares account for some of the decline. After 2000, the counterfactual series are very close to our baseline $\bar{u}_t$ estimate, implying that the majority of the secular decline after 2000 reflects the trends in the group-specific flows rather than changes in the composition.

THE ROLE OF YOUNG WORKERS AND FEMALES The bulk of the decline in the secular trend in the unemployment rate is accounted for by the group-specific trends of young workers (age 16–24) and prime age females until 2000. After the late 2000s, the decline is more broad based and goes beyond young workers and females. We conduct two counterfactual exercises:

(1) The measure $\bar{u}_t$ when the inflow rate of young and prime age females is held constant at its 1985 level (the dashed line). The secular trend in the unemployment rate would have experienced only a modest decline by 2000, highlighting the importance of females' role in the secular trend. After the late 2000s, the counterfactual series exhibits a substantial decline, pointing to a more broad-based decline.

(2) The measure $\bar{u}_t$ when, in addition to point 1, young male workers also experienced constant inflow rates at their 1985 level (the dotted line). The secular trend in the unemployment rate would have been about constant up until 2010, before declining to slightly above 6.5 percent.

THE IMPACT OF THE GREAT RECESSION Prime age males and females age 55 and older are the only two demographic groups that experienced a differential trend in their inflow rate after 2007. We interpret this change as the effect of the Great Recession and carry out a counterfactual exercise to capture its effect:

(1) The measure $\bar{u}_t$ when we fix the inflow rate for prime age males and females over 55 at their 2006 level, right before the onset of the Great Recession (the dashed line; the graph begins in 2005). The declines in the

inflow rate for these subgroups have only a modest effect on the secular trend, accounting for less than 50 basis points at maximal impact. This suggests that the declining trend in the inflow rate is primarily driven by forces in place well before the beginning of the last recession.

This set of counterfactuals illustrates clearly that most of the steady decline in the secular trend in inflow rates is associated with strong declines in the group-specific, trend inflow rate of young workers and prime age females observed from the early 1980s up to the late 2000s. The secular decline in the last two decades or so is more broadly based and cuts across demographic groups.

III. A Simple, Forward-Looking Phillips Curve

Thus far, we have focused on the secular trend in the unemployment rate implied by trends in labor market flows. This is, however, conceptually distinct from the "natural rate of unemployment," u_t^*, which is defined as the unemployment rate such that, controlling for supply shocks, inflation remains stable. Although the New Keynesian model—which has become a popular framework for monetary policy analysis and the core structure in many monetary DSGE models—features a forward-looking Phillips curve, it is silent on u_t^*. Instead of relating inflation to the unemployment gap, it typically relates inflation to the output gap or real marginal costs (Woodford 2003; Galí 2015). However, Galí (2011) reintroduces the unemployment rate into a New Keynesian model by rewriting the wage inflation equation, albeit with the assumption of a constant natural rate of unemployment. As described next, this motivates our formulation of the New Keynesian Phillips curve that connects inflation, π_t, to the unemployment gap, $u_t - u_t^*$, and retains forward-looking inflation expectations.

III.A. A Stylized New Keynesian Phillips Curve with Unemployment

We motivate our empirical specification of the New Keynesian Phillips curve with a stylized model based on the work of Galí (2011). In this framework, unemployment arises as a result of the market power of workers, which is reflected in positive wage markups. In particular, we assume a large representative household with a continuum of members specializing in different types of labor services and whose members experience different levels of disutility from working. Prices are fully flexible, but nominal wages are sticky; in each period, workers of a given type get to reset their wages with probability $1 - \theta_w$, similarly to that done by Christopher Erceg, Dale Henderson, and Andrew Levin (2000) and by Guillermo Calvo (1983).

Monopolistically competitive firms have access to a linear production function and produce using labor as the only input. Optimizing firms equate their marginal revenue and marginal costs:

$$a_t = w_t - p_t, \tag{16}$$

where the exogenous process a_t is the combination of (log) productivity and markup shocks to firms, w_t denotes the log nominal wage, and p_t is the log of the good's price. Given the wage, firms choose the quantity of workers employed, and the household assigns the workers with the lowest disutility of working. Because labor supply is elastic along the extensive margin, higher wage markups result in higher participation and therefore higher unemployment in the economy.

When they get to reset their wages, workers choose new wages that are a markup $\mu_{w,t}$ over a weighted average of current and expected future price-adjusted marginal rates of substitution. This results in a log-linearized wage Phillips curve of the form

$$\pi_t^w = -\kappa_w \left(\mu_{w,t} - \mu_{w,t}^*\right) + \beta E_t \pi_{t+1}^w, \tag{17}$$

where $\kappa_w = (1 - \theta_w)(1 - \beta\theta_w)/(\theta_w((1 + \varphi\varepsilon_w)) > 0$, φ is the steady state labor supply elasticity; $-\varepsilon_w$ is the steady state elasticity of demand for labor of different types; $\pi_t^w = w_t - w_{t-1}$ denotes nominal wage inflation; $\mu_{w,t}$ is the cross-sectional average wage markup over the economy's average marginal rate of substitution; and $\mu_{w,t}^*$ is an exogenous process capturing the time variation in the markup in the labor market, which in turn depends on firms' elasticity of demand for labor of different types, as well as that of the labor supply elasticity. Iterating equation 17 forward, we obtain

$$\pi_t^w = -\kappa_w \left(\mu_{w,t} - \mu_{w,t}^*\right) - \kappa_w E_t \sum_{T=t}^{\infty} \beta^{T-t+1} \left(\mu_{w,T+1} - \mu_{w,T+1}^*\right). \tag{18}$$

When average wage markups are below their desired level, workers who reset their wage will adjust it upward, resulting in positive wage inflation. Equation 18 reveals the central feature of the New Keynesian Phillips curve. The current gap is only one driver of inflation, and it might be a small contributor in the empirically relevant case if the slope of the curve, κ_w, is relatively flat. However, the discounted expected future path of the gap is a determinant of inflation as well. For a given level of the current gap, shifts in expectations have important implications for wage inflation—an insight lacking in the traditional backward-looking Phillips curve. The implication is then that it is important to take into account expectations when analyzing the relation between wage inflation and the markup gap.

Workers participate in the labor market only if their real wage is above their disutility from working, and Galí (2011) shows that this implies that the unemployment gap is proportional to the markup gap, so wage inflation can be expressed as

$$\pi_t^w = -\kappa x_t + \beta E_t \pi_{t+1}^w, \tag{19}$$

where $\kappa = \kappa_w \varphi > 0$, and x_t denotes the unemployment gap; here, the natural rate of unemployment (in deviation from its trend), defined as $z_t = u_t^* - \bar{u}_t = \varphi^{-1}\mu_{w,t}^*$, captures, in turn, time variation in firms' elasticity of demand for different types of labor, as well as in labor supply elasticity. Finally, using the identity relationship between price and wage inflation, we have

$$\pi_t^w = \pi_t + (w_t - p_t) - (w_{t-1} - p_{t-1}) = \pi_t + \Delta a_t, \tag{20}$$

where for the last equality, we use firms' profit-maximizing condition (equation 16). Using this expression to replace wage inflation in the wage Phillips curve, we obtain the price inflation New Keynesian Phillips curve, expressed in terms of the unemployment gap:

$$\pi_t = -\kappa x_t + \beta E_t \pi_{t+1} + \beta E_t (\Delta a_{t+1} - \Delta a_t), \tag{21}$$

where the last component is an exogenous term measuring expected wage growth, which depends on productivity and price markup shocks.

III.B. An Empirical Model

The model just described can be generalized in a variety of ways. In particular, we allow wages that are not optimally reset to be indexed to a combination of lagged inflation and the inflation target to better capture features of the data. Assuming rational expectations, the Phillips curve we consider in our empirical model thus takes the form

$$(22)\quad \pi_t - \pi_t^* = \gamma\left(\pi_{t-1} - \pi_{t-1}^*\right) - \gamma\sigma_{\pi^*}\varepsilon_t^{\pi^*} - \kappa E_t \sum\nolimits_{T=t}^{\infty} \beta^{T-t} x_T - \beta \frac{1-\rho_a}{1-\beta\rho_a} \Delta a_t .$$

Here, π_t is determined by five core components: (1) π_t^*, which represents the drift in long-term inflation expectations, and therefore the degree of anchoring, and is assumed to evolve as

$$(23)\qquad \pi_t^* = \pi_{t-1}^* + \sigma_{\pi^*}\varepsilon_t^{\pi^*};$$

(2) π_{t-1}, which captures inertia in the inflation process; (3) $x_t = u_t - u_t^*$, which denotes the current unemployment gap, which evolves as

$$(24)\qquad x_t = a_{x,1}x_{t-1} + a_{x,2}x_{t-2} + \sigma_x\varepsilon_t^x ,$$

and here we adopt the common assumption of an exogenous data-generating process for the unemployment gap (Laubach 2001; Galí 2011); (4) the discounted expectation of future unemployment gaps, discounted at the rate β; and (5) the shock Δa_t, which we assume evolves as

$$(25)\qquad \Delta a_t = \rho_a \Delta a_{t-1} + \sigma_a\varepsilon_t^a .$$

In terms of the structural model described above, this shock corresponds to real wage growth. Given the evolution of the unemployment gap, we can then rewrite equation 22 as

$$(26)\qquad \pi_t - \pi_t^* = \gamma(\pi_{t-1} - \pi_{t-1}^*) - \gamma\sigma_{\pi^*}\varepsilon_t^{\pi^*} - \kappa w_{\pi,1}x_t - \kappa w_{\pi,2}x_{t-1} + \varsigma_t ,$$

where $\varsigma_t = -\beta \dfrac{1-\rho_a}{1-\beta\rho_a}\Delta a_t$; $w_{\pi,1} = (1 - \beta(a_{x,1} + \beta a_{x,2}))^{-1}$; and $w_{\pi,2} = \beta a_{x,2} \cdot w_{\pi,1}$. As a result, observed inflation is measured as

$$(27)\qquad \prod\nolimits_t = (\pi_t - \pi_t^*) + \pi_t^* ,$$

and inflation expectations at different horizons j can be written as

$$(28)\qquad \mathbb{E}_t \prod\nolimits_{t+j} = \pi_t^* + \ell_\pi' F^j X_t ,$$

where $X_t = (\pi_t - \pi_t^*, x_t, x_{t-1}, \varsigma_t)'$, $\ell_\pi = (1, 0, 0, 0)'$ and $F = F(a_{x,1}, a_{x,2}, \gamma, \rho_a)$ is determined by equations 24 through 26.

The unemployment rate, u_t, may be expressed in terms of

$$u_t = x_t + z_t + \bar{u}_t, \tag{29}$$

with $z_t = u_t^* - \bar{u}_t$, the deviation of the natural rate of unemployment from its secular trend, which follows:

$$z_t = \rho_z z_{t-1} + \sigma_{z,\varsigma}\sigma_\varsigma \varepsilon_t^z. \tag{30}$$

This specification allows for persistent deviations of u_t^* from the secular trend, but imposes that over the longer run, these deviations shrink toward zero.

The model can be cast in state-space form. Equations 23 through 26, together with equation 30, represent the transition equations in the state-space model, and equations 27 through 29 are the observation equations. We estimate the model over the sample 1960:Q1–2018:Q3 using quarterly data. Our observed measure of u_t is the civilian unemployment rate from the Bureau of Labor Statistics (BLS). Inflation is measured as core CPI inflation in quarterly annualized percentage changes, which are also available from the BLS. We obtain a range of inflation expectations from various surveys by professional forecasters. For short-term inflation expectations, we combine six-month-ahead expectations, averaged across forecasters, from the Livingston Survey (available at semiannual frequency through our sample) and the Survey of Professional Forecasters (SPF, available since 1981:Q3). For long-term inflation expectations, we combine 5- to 10-year-ahead forecasts from Blue Chip Economic Indicators, Blue Chip Financial Forecasts, and the SPF. For the years 1975–77, we also use 5- to 10-year-ahead inflation expectations from the University of Michigan's Consumer Sentiment survey (see online supplemental appendix B for additional details about each series). Using equation 2, model-consistent, six-month-ahead inflation expectations are given by

$$\pi_t^* + (1/2)\,\ell_\pi' \sum_{j=1}^{2} F^j X_t, \tag{31}$$

while 5- to 10-year-ahead expectations can be expressed as

$$\pi_t^* + (1/20)\,\ell_\pi' (1 - F)^{-1} (1 - F^{20}) F^{20} X_t. \tag{32}$$

We include independent measurement errors for both short-term and long-term forecasts with standard deviation parameters σ_{12Q} and σ_{510Y}. All parameters are estimated using Bayesian methods, with the exception of the discount rate, β. This parameter is set to $\beta = 0.99$, a value commonly used in the literature (Woodford 2003). We split the parameters in two vectors; $\Theta^1 = (\gamma, \kappa, a_{x,1}, a_{x,2}, \rho_a, \rho_z, \sigma_{z,\varsigma})'$ and $\Theta^2 = (\sigma_x^2, \sigma_\varsigma^2, \sigma_\pi^{*2}, \sigma_{12Q}^2, \sigma_{510Y}^2)'$. Conditional on observing $\bar{u}_t$, this linear model can be estimated using the Gibbs sampler. In the first step, the Metropolis Hastings algorithm is used to draw parameters from Θ^1 for which we do not know the posterior distribution (this is due to the fact that the matrices F^j are nonlinear functions of the underlying model parameters).[12] In the second step, the Kalman smoother is used to draw the unobserved states, including initial conditions. In the third step, conditional on the drawn unobserved states, parameters from Θ^2 are drawn using known posterior distributions.[13] Because we observe a full distribution of paths of $\bar{u}_t$, we have to add an additional step in the estimation to account for this uncertainty. We first draw a path for $\bar{u}_{1:T}$, obtained from the estimation in section II, and conditional on this draw, we then proceed with the Gibbs sampler as described. We repeat this estimation procedure for a number of $\bar{u}_t$ paths selected at random.[14] This approach is motivated by the assumption, discussed in section II, that the unemployment trend $\bar{u}_t$ is exogenous to the variables in the Phillips curve model.

III.C. Adding Information from Wages

Although subsection III.A characterizes a simple wage Phillips curve, the discussion so far has not characterized the information available from observed wages. The importance of wages for assessing the unemployment gap has been emphasized by, for example, Robert Solow (1964), Olivier Blanchard and Peter Diamond (1989), and Galí (2011). Here we consider a second specification that includes *both* price and wage inflation. The goal here is to evaluate the impact of this additional information on our estimates of the natural rate of unemployment. Given that wages are measured with a considerable degree of noise, we extract information from five alternative data sources.[15] From the Employment Cost Index

12. For a comprehensive review of these Bayesian methods, see Herbst and Schorfheide (2015).

13. The estimation method follows quite closely del Negro and others (2017), who provide full details.

14. In more detail, we draw 250 paths from the distribution of $\bar{u}_t$. For each of these draws, we run a chain of 10,000 draws of the model parameters and states with the Gibbs sampler. Importantly, each parameter's chain is initialized by looking for a set of parameters close to the mode. We then randomly select 250,000 draws that we use to compute the joint distribution of parameters and states.

15. We thank our discussant, Steven Davis, for this suggestion.

release, we use growth in wages and salaries for private industry workers, along with growth in total compensation for all civilian workers (both starting in 2001:Q1). From the Establishment Survey, as part of the Employment Situation release, we use growth in average hourly earnings of all private sector employees and growth in average hourly earnings of production and nonsupervisory employees (starting in 2006:Q1 and 1964:Q1, respectively). From the Productivity & Costs release, we use growth in compensation per hour of the nonfarm business sector (starting in 1947:Q1). All data are available from the BLS, and growth rates are expressed at a quarterly, annualized rate.

The relation between wage and price inflation implied by the model in equation 20 implies

$$\begin{aligned} \pi_t^w = {} & g_w + \pi_t^* + \gamma(\pi_{t-1} - \pi_{t-1}^*) - \gamma\sigma_{\pi^*}\varepsilon_t^{\pi^*} \\ & - \kappa w_{\pi,1} x_t - \kappa w_{\pi,2} x_{t-1} - \frac{\beta^{-1} - 1}{1 - \rho_a}\varsigma_t, \end{aligned} \tag{33}$$

where π_t^w denotes nominal wage growth and g_w is the (constant) mean growth rate of real wages. This can be used to obtain this additional measurement equation to the model

$$\prod_t^{w,(j)} = \Theta^{(j)}\pi_t^w + \xi_t^{(j)}$$

where $j = 1, \ldots, 5$ denotes the individual nominal wage growth measures introduced above and where $\xi_t^{(j)}$ are first-order autoregressive measurement errors with autocorrelation coefficient $\rho_{\xi,(j)}$ and innovation standard deviation $\sigma_{\pi_{w,(j)}}$. We normalize the first loading coefficient $\Theta^{(1)} = 1$ and then estimate the remaining loadings $(\Theta^{(2)}, \ldots, \Theta^{(5)})$ along with $(g_w, \rho_{\xi,(1)}, \ldots, \rho_{\xi,(5)}, \sigma_{\pi_{w,(1)}}, \ldots, \sigma_{\pi_{w,(5)}})$. We view this as a particularly compelling exercise, given the relative stability of the wage Phillips curve over the sample, as shown by Galí (2011). In fact, the equation given above has a form similar to the one estimated by Galí (2011)—with a few key differences. Our specification includes an inflation trend, π_t^*, and, importantly, the original specification assumes a constant level of the natural rate of unemployment, which we eschew. We estimate this equation jointly with the rest of the model described in subsection III.B.

Table 1 shows the assumptions on the priors along with the properties of the posterior distribution for both model specifications. Notice first that the

Table 1. Parameter Estimates[a]

	Prior			Posterior (inflation only)			Posterior (inflation and wage inflation)		
Variable	*Distribution*	*Mean*	*Standard deviation*	*Mean*	*5%*	*95%*	*Mean*	*5%*	*95%*
$a_{x,1}$	Gamma	1.25	0.200	1.59	1.50	1.68	1.53	1.45	1.61
$a_{x,2}$	Normal	0.000	1.00	–0.618	–0.703	–0.529	–0.568	–0.639	–0.491
θ_w	Beta	0.750	0.100	0.879	0.847	0.906	0.852	0.820	0.878
κ	—	0.113	0.114	0.019	0.011	0.031	0.028	0.018	0.041
γ	Beta	0.500	0.265	0.049	0.004	0.196	0.025	0.003	0.060
ρ_{u^*}	Beta	0.950	0.035	0.982	0.961	0.994	0.980	0.962	0.992
ρ_a	Beta	0.500	0.200	0.366	0.204	0.449	0.350	0.279	0.411
σ^2_x	InvGamma	0.12	—	0.043	0.032	0.054	0.043	0.032	0.055
σ^2_ς	InvGamma	1.00	—	1.44	1.23	1.68	1.34	1.13	1.56
$\sigma^2_{\pi^*}$	InvGamma	0.12	—	0.067	0.053	0.083	0.069	0.054	0.087
$\sigma_{z,\varsigma}$	InvGamma	0.150	0.050	0.103	0.077	0.132	0.120	0.090	0.152
$\sigma^2_{o,\pi^{510Y}}$	InvGamma	0.005	—	0.003	0.001	0.007	0.004	0.001	0.010
$\sigma^2_{o,\pi^{12Q}}$	InvGamma	0.500	—	0.153	0.122	0.188	0.165	0.128	0.208
g_w	Normal	0.4	0.05	—	—	—	0.439	0.363	0.521

Source: Authors' calculations.

a. The variable κ is calculated assuming $\kappa = (1 - \theta_w)(1 - \beta\theta_w)/\theta_w$.

priors for the innovations' variance report only the mean, as the standard deviation is not defined; for these priors, we use an Inverse-Gamma distribution with shape parameter of 1.5, enough to have a well-defined mean. Second, the posterior distribution of the parameters for the two model specifications is broadly similar, with a few small differences that are discussed below. The process for the unemployment gap shows a high degree of persistence, consistent with medium-frequency business cycle movements in the unemployment rate. Regarding the Phillips curve, the slope is precisely estimated and in the range of about 0.02–0.03 across specifications, and it implies a fairly flat curve, as is often found in the literature (del Negro, Giannoni, and Schorfheide 2015). The addition of wages as an observable delivers a slightly steeper slope, but it does not fundamentally alter the estimated link between the current gap and inflation. The Phillips curve does not display significant inflation inertia, given the estimate of γ in the range about 0–0.2. The process for z_t is estimated to be highly persistent consistent with prolonged deviations of u_t^* from the secular trend in unemployment. The signal-to-noise ratio, $\sigma_{z,\varsigma}$ is tightly estimated in the range of about 0.12–0.15. Also, in this case, the introduction of wage inflation implies a somewhat more volatile natural rate of unemployment, as we discuss in the next section. Finally, the measurement errors for the survey-based measures of expectations are estimated to have small variances, allowing a tight mapping from observed expectations to the unobserved unemployment gap.

Priors for the five first-order autoregressive measurement errors (not shown in table 1) in the wage equation are as follows. The prior on the innovations corresponds to an inverse-gamma with shape and scale parameters implying a diffuse prior consistent with a well-defined mean of 1. The prior on the autocorrelation coefficient is normally distributed, with zero mean and variance determined using the same parameterization, with $\lambda = 0.1$ as in the VAR Minnesota prior. The priors are fairly diffuse. Finally, the priors on the loadings $\Theta^{(2)} \ldots \Theta^{(5)}$ (also not shown in the table) are defined as independent normal densities with mean 1 and standard deviation 0.5.

Before moving on to the main empirical results, it is useful to consider a simplified version of the model in order to make concrete the appropriate interpretation of κ, the slope of the Phillips curve, in this forward-looking model. As mentioned above, the estimated slope is fairly small; however, this does not necessarily imply a weak link between the unemployment gap and inflation. For example, consider a simpler model, where the unemployment gap, x_t, is a first-order autoregressive process (that is, $a_{x,2} = 0$) and set $\gamma = 0$. Then, solving forward for expectations delivers this relation between inflation and the contemporaneous unemployment gap,

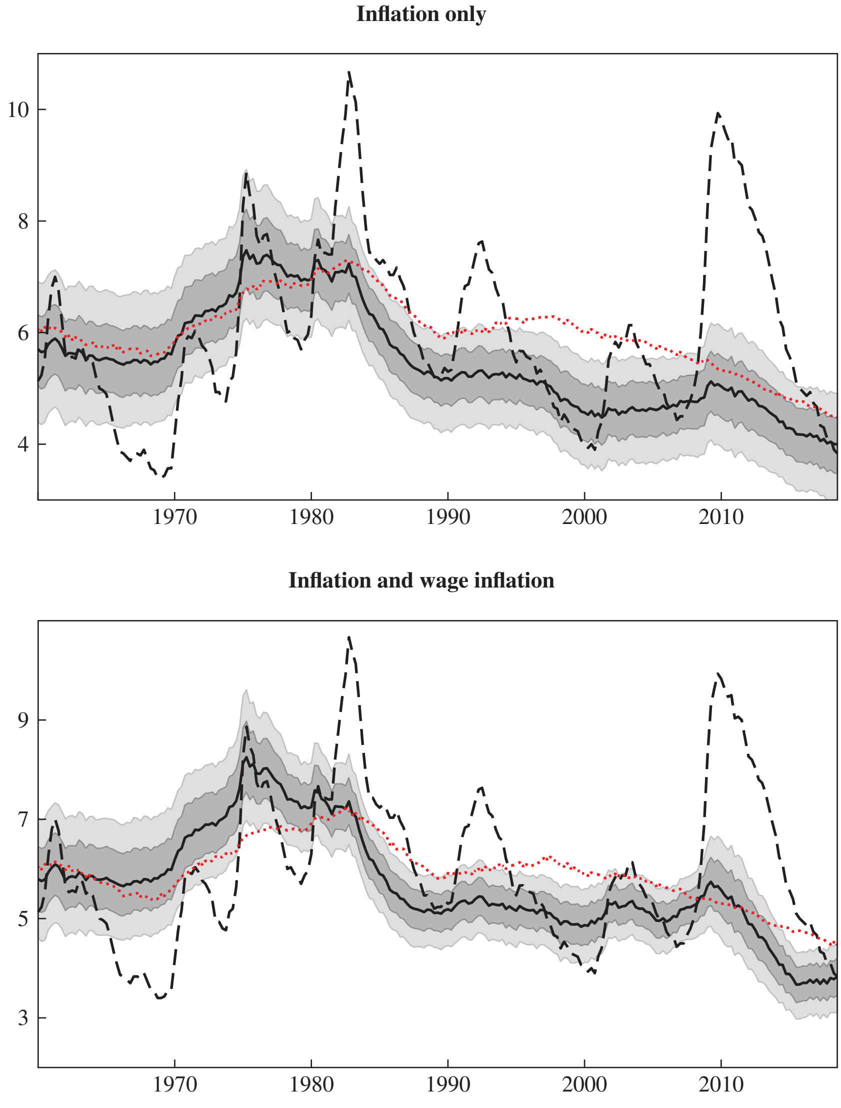
$$\pi_t - \pi_t^* = -\frac{\kappa}{1 - a_{x,1}\beta} x_t + \varsigma_t .$$

Given that $\beta = 0.99$ and that we would expect a reasonably high degree of persistence in the unemployment gap, so that $a_{x,1}$ is near 1, then the coefficient relating actual inflation (in deviation from the trend) and the unemployment gap would be much larger than κ. The same intuition applies to the more general model introduced in this section.

IV. Measuring u_t^*

We estimate u_t^* since 1960 and show its evolution in the two panels of figure 8. The top panel refers to the model using price inflation only, and the bottom panel shows the results where both prices and wages are used. Both model specifications yield comparable predictions in general, with some differences, which are highlighted below. Overall, the natural rate of unemployment is estimated quite precisely with a 95 percent confidence interval of about 2 percentage points in the model using price inflation only, and with even narrower bands when wage information is added. When discussing ranges of the possible values of u_t^* at any particular point in time, we refer to the 68 percent interval. In the first decade of the sample, the natural rate hovers slightly below 6 percent and starts rising in the early 1970s, reaching comfortably above 7 percent by the late 1970s before falling to about 7 percent in 1983. The increase in the natural rate was the subject of a heated debate during the 1970s. Going back to earlier papers—such as those by Hall (1970a, 1970b), Gordon (1972, 1982), Perry (1978), and James Tobin (1974)—there appears to have been a consensus that the natural rate of unemployment increased to somewhere between a low of 5.0 percent and a high of 7.0 percent. Interestingly, these insightful analyses did not get much traction in policy circles, and the Humphrey-Hawkins Full Employment and Balanced Growth Act of 1978 set an unemployment target of 4 percent for 1983. Subsequent research devoted substantial effort to understanding this period. For example, Summers (1986, 340) states that "where Kennedy-Johnson economists set 4 percent as an interim full-employment target, contemporary policymakers would regard even the temporary achievement of 6 percent unemployment as a great success." The natural rate then declines throughout the 1980s, consistently below

Figure 8. The Natural Rate of Unemployment, u_t^*, 1960–2018[a]



Source: Authors' calculations.

a. This figure shows the estimate of u_t^* for the inflation-only specification (top panel) and the inflation and wage inflation specification (bottom panel). The dotted line denotes the median $\bar{u}_t$. Shading denotes 68 percent and 95 percent confidence intervals.

the median secular trend in the unemployment rate (the dotted line). More recent analysis by Laurence Ball and N. Gregory Mankiw (2002) estimated the natural rate to be about 5.4 percent in 1960, and rising to 6.8 percent in 1979 and decreasing to 4.9 percent in 2000. Douglas Staiger, James Stock, and Mark Watson (1997) also have similar estimates.

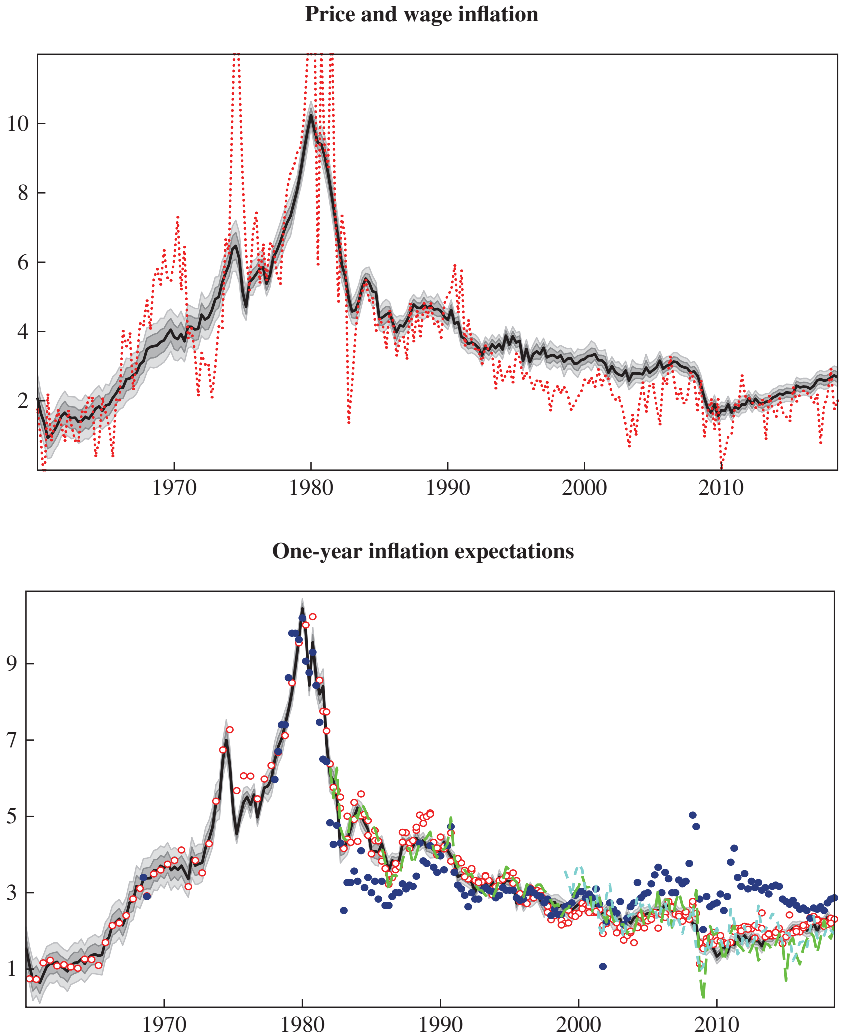
One of the key differences between the top and bottom panels concerns the behavior of the natural unemployment rates during the 1970s. While the model with prices only estimates the natural rate to increase along with the secular trend in unemployment, the richer specification including wages estimates a further increase in the natural rate. One possible explanation for this discrepancy is the wage-price controls implemented in the early 1970s and their relative effects on wage growth and price inflation. As shown in the top panel of figure 9, while inflation dips in the early 1970s, wage inflation remains robust, signaling a strongly negative unemployment gap. The Nixon administration imposed wage and price controls in August 1971 that lasted until April 1974. The control program went through four phases. The first two phases were more strict and accomplished only a slight reduction in wage growth but a marked decline in the rise in prices between 1971:Q3 and 1972:Q2.[16] Phase II (which lasted until January 11, 1973) was followed by phases III and IV, but controls were generally relaxed in the last phases. Inflation started picking up in late 1972 while wage growth moderated. By the time wage-price controls were dismantled in April 1974, U.S. inflation had reached double digits. In fact, both panels of figure 10 show that, regardless of model specification, a substantial negative unemployment gap remained until the early 1980s.

The period spanning the 1990s to the Great Recession was characterized by a fairly stable natural rate of unemployment, which remained range-bound between 4.5 and 5.5 percent. During this period, the median u_t^* remained consistently below its secular trend. To speculate, some of this decline might have been due to the rapid growth of technological progress during the period.[17] As shown in figure 9, the unemployment gap had been consistently positive throughout the 1980s, around the deep monetary contractions of the Volcker disinflation period. It turned negative briefly in

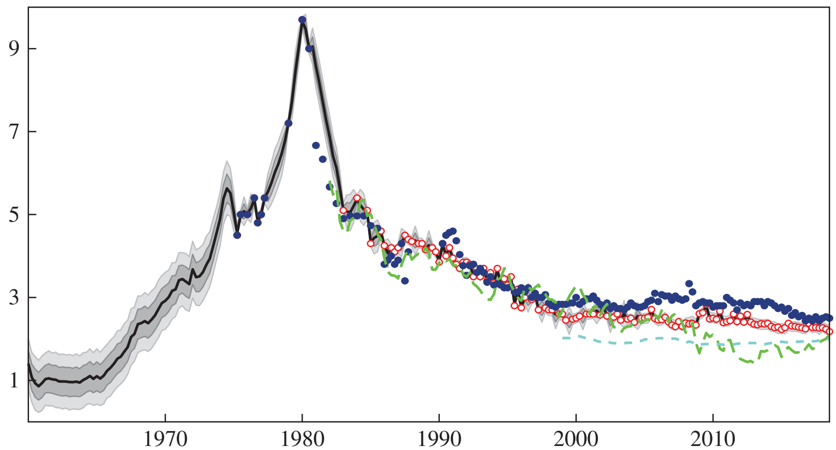
16. See Gordon (1972, 1973).

17. As then–chair of the Federal Reserve Alan Greenspan said in a speech in 1998 on the New Economy: “Coupled with the quickened pace of productivity growth, wage and benefit moderation has kept growth in unit labor costs subdued in the current expansion. This has both damped inflation and allowed profit margins to reach high levels” (Greenspan 1998).

Figure 9. Inflation, Inflation Expectations, and Wages, 1960–2018[a]



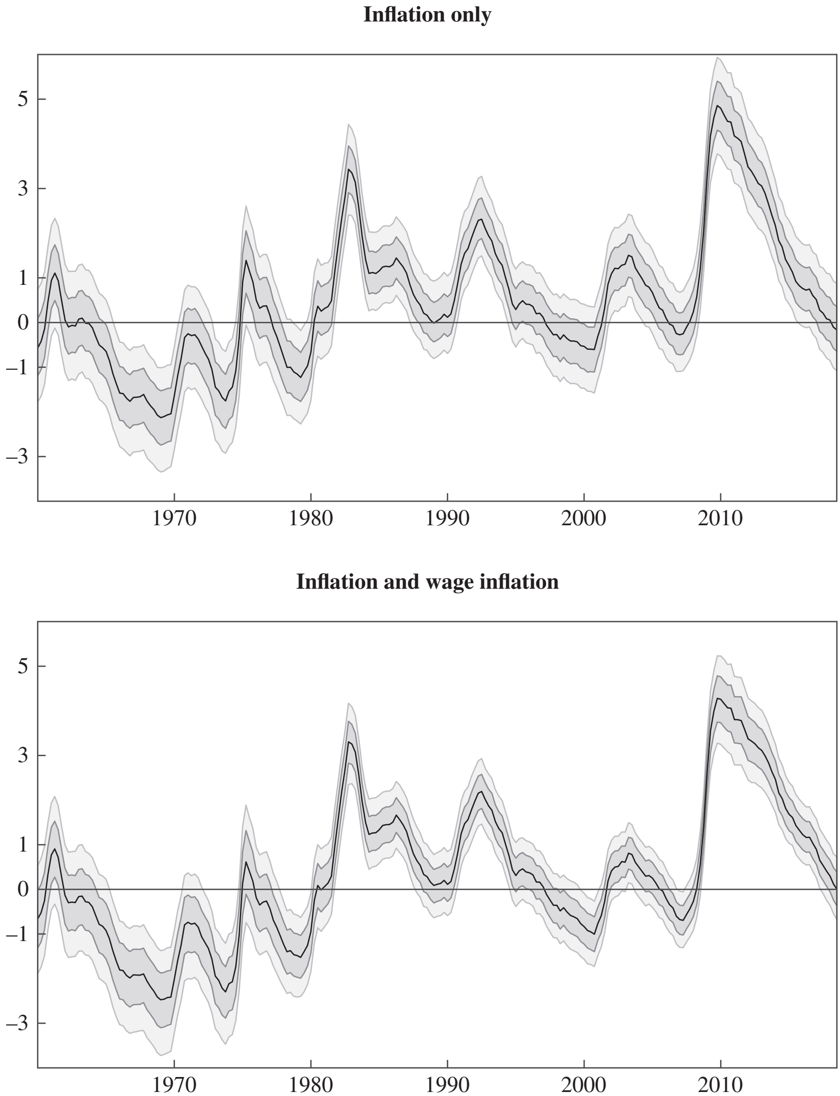
Long-term inflation expectations



Sources: Bureau of Labor Statistics; Survey of Professional Forecasters; Livingston Survey; Blue Chip Economic Indicators; Blue Chip Financial Forecasts; University of Michigan; Abrahams and others (2016); Haubrich, Pennacchi, and Ritchken (2012).

a. The top panel shows realized quarterly annualized inflation (the dotted line) and the model predicted quarterly nominal wage inflation distribution (the solid black line and gray shading). The middle panel shows survey-based one-year inflation expectations of professional forecasters (the hollow dots) and households (the solid dots), and model-implied expectations (black line and gray shading). The short- and long-dashed lines show inflation expectations extracted from market prices from Abrahams and others (2016) (2-year) and Haubrich, Pennacchi, and Ritchken (2012) (one year), respectively. The bottom panel shows the same series from the middle panel but for the 5-year horizon beginning in 5 years. Shading denotes 68 percent and 95 percent confidence intervals.

Figure 10. The Unemployment Gap, 1960–2018[a]



Source: Authors' calculations.

a. This figure shows the estimated unemployment gap, $u_t - u_t^*$, for the inflation-only specification (top panel) and the inflation and wage inflation specification (bottom panel). Shading denotes 68 percent and 95 percent confidence intervals.

the late 1990s, but this dip was preceded and followed by the 1990–91 and 2001 recessions.

Finally, during the prerecession years 2005–6, the natural rate of unemployment began increasing toward its long-run trend. This period presents the second important difference between the two model specifications. Including both prices and wages in the estimation leads to a higher estimate of u_t^*, which ends up overshooting its long-run trend, with its median estimate peaking in 2009–10 at just under 6.0 percent. Conversely, the model specification employing only price inflation predicts a milder increase (with a median that peaks at about 5.1 percent).[18] The different estimates reflect a pickup in wage growth in the period 2005–6, which we do not see in the measure of price inflation. Subsection IV.B. discusses the possible driving forces behind this increase. In the aftermath of the Great Recession, the natural rate of unemployment gradually declined, roughly in line with its secular trend. This finding implies that the fear of hysteresis after the Great Recession did not materialize, as we discuss in the next subsection. Both model specifications deliver estimates of the natural rate toward the end of 2018 in the range of 3.5 to 4.5 percent, which is consistent with the current unemployment gap of about zero.

Importantly, as shown in figures 9 and 10, the estimated Phillips curve is consistent with periods of large slack in the labor market and relatively stable inflation. This is perfectly illustrated by the Great Recession, which displays the largest unemployment gap in the sample, of about 4 percentage points, while price inflation declined only modestly. Indeed, while core CPI inflation was averaging 2.3 percent from 2005 to 2007, it declined to 1.4 percent on average from 2009 to 2011. Most important for the stability of inflation is the fact that inflation expectations declined only modestly both during and after the Great Recession, as shown in figure 9. As indicated in our Phillips curve, equation 22, inflation expectations reflect the expected path of future unemployment gaps, and so the near-stability of inflation expectations in the aftermath of the Great Recession suggests that the unemployment gap was expected to close. This is consistent with the attenuated response of inflation to the large unemployment gap.

Our analysis of the Great Recession, through the lens of our estimation results, does not, however, imply that inflation is necessarily insensitive to the unemployment gap. In fact, we see that a somewhat smaller rise in the

18. This is in line with the estimates by Daly and others (2012), who estimated the natural rate to be between 5.5 and 6.6 percent, with a preferred estimate at 6 percent at the end of 2011.

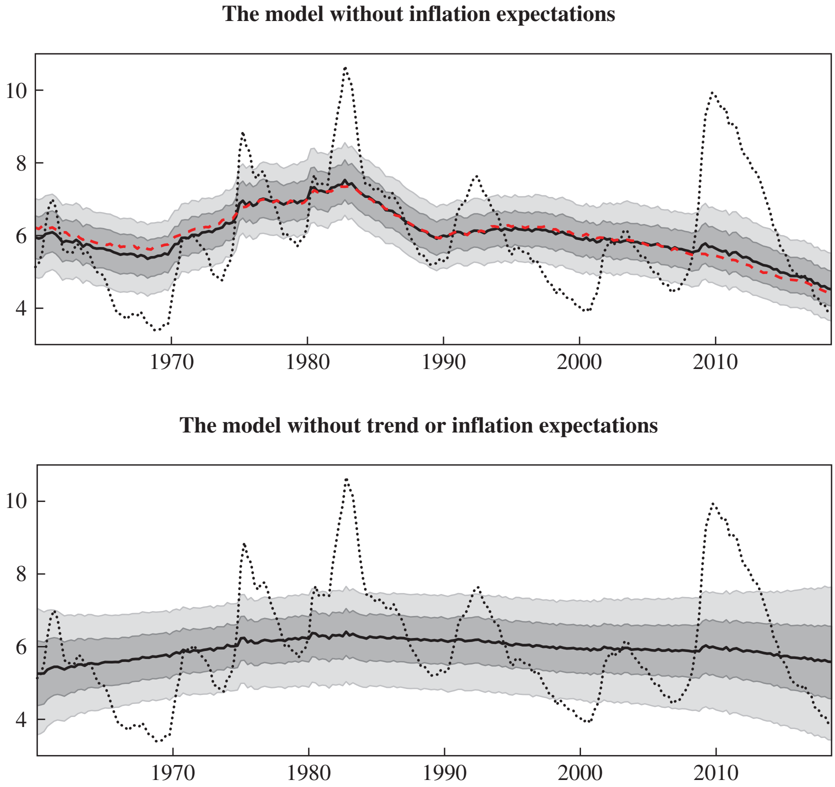
unemployment gap in the early 1980s caused a much more significant drop in inflation, with average core CPI inflation falling from 10.9 percent in the 1979–81 period to 4.5 percent in 1983–84. The key determinant is the behavior of inflation expectations, which dropped much more sharply in the early 1980s than was the case after the Great Recession. A comparison of the early 1980s with the Great Recession period stresses the importance of accounting for inflation expectations in explaining the behavior of inflation and the unemployment gap, and hence for estimating u_t^*.

The middle panel of figure 9 shows the model-implied predictive distributions (the gray shaded area) for the one-year-ahead inflation forecast, together with measured expectations from professional forecasters (the hollow dots). It is worth pointing out that as inflation and inflation expectations have been reverting to their long-term trend (see the figure's bottom panel), the unemployment gap has been steadily closing. The figure's middle panel also shows that alternative measures of one-year-ahead expectations display a roughly similar pattern as those of professional forecasters. For example, measures of inflation expectations extracted from asset prices (the dashed lines) are broadly in line. The median of households' expectations from the Michigan survey (the solid dots) behave differently since the early 2000s, predicting considerably higher inflation. Olivier Coibion and Yuriy Gorodnichenko (2015) show how the difference between the findings of household and professional forecasters over this period can be explained by oil prices. This difference, however, shrinks significantly when one looks at long-term inflation expectations, shown in the figure's bottom panel. These measures are of particular importance because they show the degree of anchoring of inflation expectations. As can be gleaned from the figure, all measures display a stable pattern after 1998, albeit at different levels, providing additional evidence to why the large unemployment gaps over these years were not associated with deflation.

IV.A. The Information Content of Inflation Expectations and the Secular Trend in the Unemployment Rate

In this subsection, we assess the role of observed inflation expectations and the secular trend in the unemployment rate in our estimated u_t^*. Figure 11 shows the results from two different estimation exercises. The top panel shows the estimate of u_t^* (for the price inflation only specification) when only information about $\overline{u}_t$ is provided, along with the realized unemployment rate and inflation rate. In this case, the estimated u_t^* is essentially identical to $\overline{u}_t$ (the dashed line), emphasizing the key role

Figure 11. Phillips Curve Models without Key Inputs, 1960–2018[a]



Source: Authors' calculations.

a. This figure shows results from an inflation-only Phillips curve model without key inputs. The top panel shows the estimated u_t^* (the black line) without inflation expectations as inputs. The dashed line represents the median estimate of $\bar{u}_t$. The bottom panel shows the estimated u_t^* (the black line) without information from inflation expectations or the secular trend in the unemployment rate. Shaded areas denote 68 percent and 95 percent confidence intervals.

that inflation expectations play in identifying movements in the unemployment gap across the state of the business cycle (the same result is obtained including both price and wage inflation in the estimation). In the bottom panel, we show the resulting u_t^* estimates when we further remove the secular trend from the observables. For this specification, we supply a stochastic process for the evolution of the trend, following the well-known model developed by Thomas Laubach (2001). The natural rate of unemployment follows the process

$$\bar{u}_t = \bar{u}_{t-1} + g_t \text{ and} \tag{34}$$

$$g_t = g_{t-1} + \sigma_g \varepsilon_t^g. \tag{35}$$

The model features no forward-looking behavior. Setting $\beta = 0$ delivers the Phillips curve

$$\pi_t - (1 - \gamma)\pi_t^* - \gamma\pi_{t-1} = -\kappa x_t + \varsigma_t, \tag{36}$$

where ς_t is assumed to be independent and identically distributed under this specification. We also fix the standard deviation σ_g to deliver a smooth estimate of the trend, similar to our earlier analysis. In doing so, we considerably reduce the estimation uncertainty. In particular, we set $\sigma_g = 0.02 \times \sigma_u^*$. As is clear from the bottom panel of figure 11, the exercise shows that there is very little information about the natural rate of unemployment once we focus only on the joint behavior of inflation and unemployment. We conclude that inflation expectations and the secular trend in unemployment are therefore critical for assessing u_t^*.

IV.B. The Great Recession and Factors Affecting Matching Efficiency

It is of special interest to focus on the behavior of the natural rate of unemployment during the Great Recession and its aftermath. The Great Recession was not only the deepest postwar downturn in the labor market; it was also followed by an unprecedented period of high unemployment rates. The unemployment rate remained stubbornly high, printing at about 9 percent in January 2011, while many measures of economic activity had recovered by then. This disconnect triggered increased disagreement about the nature of the rise in the unemployment rate and whether the recession permanently affected the workings of the labor market. For example, figure C.3 in the online supplemental appendix summarizes the Federal Reserve Board's and Federal Reserve Banks' estimates of the NAIRU for three different periods: before the financial crisis in 2007, the current period (at the time), and 2015, as well as the increase between the first two periods as of January 25–26, 2011.[19] The figure shows that in 2011, there was increased disagreement not only about the current level of the natural rate but also about its level through 2015, suggesting that some participants viewed the natural rate of unemployment as higher even in the medium run due to hysteresis, as described by Blanchard and Summers (1986).

A careful examination of worker flows into and out of unemployment has shown that, though the inflow rate quickly returned to its prerecession level and gradually trended down, the persistently low outflow rate accounted for the high unemployment rate. Therefore, various explanations were suggested in the applied macroeconomics literature that operated through a long-lasting decline in the outflow rate, such as a rising mismatch, declining recruiting intensity, and a declining search effort by unemployed workers. This literature relied on rich micro data from various surveys, administrative data sources, and online data sources to quantify the effect of these factors. We next provide a simple framework derived from the search and matching literature to summarize these measures and then compare and contrast them with our measure of the natural rate of unemployment.

The point of departure is the matching function that characterizes the technology that firms and workers match with each other, building on the research of Diamond (1981), and of Dale Mortensen and Christopher Pissarides (1994), and on the work of Blanchard and Diamond (1989), who argue that changes in matching efficiency that shift the Beveridge curve may shed light on the Phillips curve. In its basic form, the inputs to the matching function at time t are the v_t vacancies posted by firms looking to hire and u_t unemployed workers looking for jobs. To accommodate the intensity of the recruiting and search effort, we denote the *recruiting intensity* of firms as q_t and the *search intensity* of workers as ι_t. A generalized Cobb–Douglas matching function that allows for shifts along the intensive margins of the firm and worker search effort can then be written as

$$h_t = \Phi_t (q_t v_t)^{\alpha} (\iota_t u_t)^{1-\alpha}, \tag{37}$$

where h_t is the total hires and $\alpha \in (0, 1)$ is the vacancy share. The term Φ_t is the aggregate matching efficiency parameter. As the specification shows,

19. See FOMC (2011a).

changes in q_t and ι_t would show up as a decline in the measured match efficiency. In addition, the mismatch between vacant jobs and unemployed workers (idle workers seeking employment in sectors, occupations, or industries different from those where the available jobs are) would manifest itself as a decline in Φ_t. Such a misalignment between the distribution of vacancies and unemployment along with a decline in the recruiting intensity and search effort would lower the aggregate outflow rate, which is defined as $f_t = h_t/u_t$.

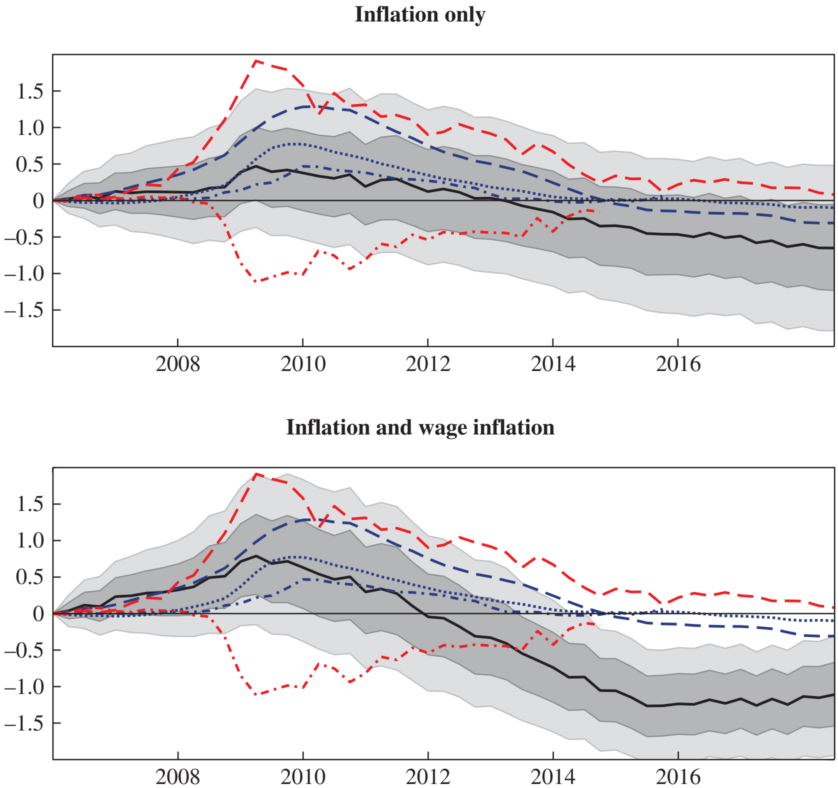
MISMATCH Şahin and others (2014) formalize the notion of mismatch by defining the economy as a large number of distinct labor markets segmented by industry, occupation, and geography.[20] Each labor market i is frictional—that is, the hiring process within a labor market is governed by a matching function. To assess the existence of mismatch, they examine whether, given the distribution of vacancies observed in the economy, it would be feasible to reallocate unemployed workers across markets in a way that reduces the aggregate unemployment rate. This involves comparing the actual allocation of unemployed workers across labor markets with an optimal allocation that assumes costless worker mobility across these markets. Because the only frictions in such an environment are the ones embodied within each market-specific matching function, unemployment arising in this environment is purely frictional. The difference in unemployment between the observed allocation and the allocation implied by the optimal environment provides an estimate of the effect of mismatch.

Şahin and others (2014) calculate mismatch unemployment at the industry level using vacancy data from the Job Openings and Labor Turnover Survey (known as JOLTS), which provides survey-based measures of job openings and hires at a monthly frequency, starting from December 2000, and at the occupation level using vacancy data from the Help Wanted Online (known as HWOL) data set provided by the Conference Board. We plot an update of their occupation and industry mismatch unemployment measures in figure 12.

RECRUITING INTENSITY Recent research by Davis, Jason Faberman, and Haltiwanger (2013) has stressed the importance of channels other than a vacancy posting in the search and matching process. They argue that channels that affect how quickly firms fill those vacancies should be taken into account as determinants of the hiring process. A variety of factors—such as

20. In online supplemental appendix A, we present a simplified version of the derivation done by Şahin and others (2014).

Figure 12. A Comparison of u_t^* with Micro-Data-Based Estimates, 2006–18[a]



Sources: Current Population Survey; Job Openings and Labor Turnover Survey; American Time Use Survey; Conference Board.

a. Estimated u_t^*, occupation and industry mismatch unemployment rates (the lowest dashed line and dotted line), the recruiting-intensity-adjusted unemployment rate (the highest dashed line), the search-effort adjusted unemployment rate (the lowest dash-dotted line), and the unemployment insurance extension adjusted unemployment rate (the highest dash-dotted line). All series show changes relative to their level in 2006:Q1. Shading denotes 68 percent and 95 percent confidence intervals.

variations in hiring standards, wages offered that differ from those of competitors, variations in the amount of screening effort, and the propensity to use informal hiring methods—all contribute to what these authors refer to as *recruiting intensity.* They generate an aggregate time series of their measure of recruiting intensity using a generalized version of a standard matching function and their derivation of the monthly evolution of hiring and vacancies in the JOLTS data. Davis, Faberman, and Haltiwanger's (2013) recruiting intensity index provides us a normalized measure of q_t in the generalized Cobb–Douglas matching function.[21] In figure 12, we plot the counterfactual unemployment rate that is computed as the difference between the actual unemployment rate and a counterfactual unemployment rate that holds recruiting intensity constant at its mean value over the sample period replicated from the "Reserve Bank Report on Structural Unemployment" by Faberman and Şahin (2011).[22] This difference reflects the effect of changes in recruiting intensity on the unemployment rate.

THE WORKER SEARCH EFFORT AND EXTENSION OF UNEMPLOYMENT INSURANCE BENEFITS Another margin that is likely to be affected by aggregate conditions is unemployed workers' search effort. One often-discussed policy that is linked to the worker search effort is the extension of unemployment insurance (UI) benefits. In theory, receiving UI benefits for a longer period reduces the incentive of the unemployed to look for work. Similarly, it also increases unemployed workers' reservation wage, so that they may reject job offers that they would otherwise have accepted in the absence of these extended benefits. During the Great Recession, unemployment insurance benefits were extended to record lengths, with individuals in most states being eligible for up to 99 weeks of UI (and, at a minimum, 60 weeks). The Federal Reserve Board's Tealbook estimated that the extension of benefits raised the natural rate of unemployment, which we plot in figure 12.[23]

However, extension of UI is not the only channel that affected the worker search effort. Toshihiko Mukoyama, Christina Patterson, and Şahin (2018) showed that during the Great Recession, the unemployment pool shifted toward workers who are more attached to the labor force, who typically search harder for jobs. They showed that, as a consequence of this shift in the composition of the unemployed as well as the increased search

21. We thank Steven Davis for providing updated data of the recruiting intensity index and Jason Faberman for sharing his replication code.

22. See FOMC (2011b).

23. See FOMC (2012).

effort in response to declining household wealth, the aggregate search effort in the economy increased during the Great Recession. They find that the increase in search intensity during and after the Great Recession moderated the increase in the unemployment rate. Absent this increase, the unemployment rate would have peaked at about 11 percent and would have been consistently higher by about 0.5 to 1 percentage point during the recovery. Figure 12 shows the effect of the rise in workers' search effort on the unemployment rate.

IV.C. The Evolution of u* and Factors Affecting the Matching Efficiency

In figure 12, we show the evolution of our estimated u_t^*, along with empirical estimates of the effects on unemployment of the persistent factors discussed in this section. All series are plotted in deviation from their level in 2006:Q1. There are two important features of these factors: (1) These persistent factors can only be measured as counterfactual gaps relative to an unemployment rate without additional assumptions; and (2) these measures are not additive, as they cannot be considered as independent from each other. That said, looking at the predictive distributions, we first observe that our measure of u_t^* aligns surprisingly well in terms of timing and magnitude with the evolution of these factors. The natural rate estimated using prices only is more aligned with industry mismatch unemployment and the effects of the extension of unemployment benefits (the dotted and dash-dotted lines in the figure). Conversely, u_t^* measured including both prices and wages in the estimation initially, is more aligned with the rise in occupational mismatch and decline in recruiting intensity. Though it displays a stronger increase relative to the inflation-only estimate, it still falls short of the observed spikes in these two series. Finally, job search intensity moved sharply, but in the opposite direction, moderating some of the effects of other factors. We also notice that these factors return to their 2006 levels in about 2014, while u_t^* continues to decline in line with the falling secular trend $\bar{u}_t$. Put differently, though the effect of the Great Recession on the unemployment rate persisted for almost a decade, these factors did normalize eventually. As such, we focus on driving forces that predate the Great Recession when studying the secular trend in the unemployment rate in the next section.[24]

24. This finding does not preclude persistent effects of the Great Recession on worker career paths (Davis and von Wachter 2011).

V. Changes in the U.S. Labor Market and Flow Dynamics

We have shown that the behavior of unemployment flows, especially the ongoing downward trend in the inflow rate, is the driver of the low levels of the natural rate of unemployment that the U.S. economy has been experiencing. In this section, we identify three important changes in the structure of the U.S. economy as the main drivers of the downward trend in the inflow rate: an increase in labor force attachment of females, a decline in job destruction and job reallocation, and the dual aging of workers and firms.

As a prelude to our analysis in investigating the economic changes that affected the evolution of the inflow rate, we report the changes in the inflow rate for the 1976–96 and 1996–2018 periods and decompose the total changes into changes accounted for by each gender and age group for the inflow rate

$$s_t \approx \sum_i \omega^s_{i,t} s^i_t, \tag{38}$$

where group i is defined as the interaction of gender and age. We consider three age groups for each gender, workers between 16 and 24 years, workers between 25 and 54, and workers older than 55 years. Table 2 decomposes the change in the inflow rate into changes accounted for by each demographic group using this simple approximation:

$$\Delta s(t,t') \approx \sum_i \left(\omega^s_{i,t'} s^i_{t'} - \omega^s_{i,t} s^i_t\right). \tag{39}$$

The table shows that females account for the majority of the decline in the inflow rate in the 1976–96 period, which coincides with the dramatic rise in the female labor force participation rate. In addition, during this period the baby boom cohort proceeded from younger ages to prime ages, reducing the aggregate inflow rate. Interestingly, the decline in the 1996–2018 period is very similar among males and females, suggesting a common factor after 1996.[25]

We also calculate the counterfactual contribution of each group, fixing their weights at their 1976 levels, a calculation that is often referred to as a shift-share analysis:

$$\Delta^C s(t,t') \approx \sum_i \left(\omega^s_{i,1976} s^i_{t'} - \omega^s_{i,1976} s^i_t\right). \tag{40}$$

25. The changes in the outflow rate are harder to interpret since the outflow rate is persistent and strongly procyclical. We report the corresponding changes in f in the online supplemental appendix.

Table 2. Shift Share with Fixed Demographic Composition[a]

		Females			*Males*		
Period	*Aggregate change*	*16–24 years*	*25–54 years*	*55+ years*	*16–24 years*	*25–54 years*	*55+ years*
A. Actual contribution							
1976–96	–0.80	–0.55	–0.24	–0.04	–0.17	0.13	0.01
1996–2018	–1.24	–0.34	–0.33	0.00	–0.34	–0.20	0.03
1976–2018	–2.04	–0.90	–0.57	–0.03	–0.51	–0.06	0.05
B. Counterfactual contribution with weights fixed in 1976							
1976–96	–0.36	–0.20	–0.32	–0.03	0.12	0.05	0.02
1996–2018	–1.29	–0.45	–0.23	–0.05	–0.42	–0.12	–0.02
1976–2018	–1.63	–0.65	–0.55	–0.08	–0.29	–0.06	0

Source: Current Population Survey.

a. Inflow rate changes are for 1976–96 and 1996–2018, and for the full sample, 1976–2018; and the actual and counterfactual contribution of each demographic group to the changes in the aggregate inflow rate. The counterfactual contributions are calculated using weights fixed at 1976 averages.

This counterfactual calculation shows that about half the decline in the 1976–96 period can be accounted for by the changing weights. In other words, if the shares of each age and gender group remained at their 1976 level, the inflow rate would have only declined by 0.36 percentage point, about 50 percent of the actual decline. The message is very different for the 1996–2018 period. Changes in the age and gender composition played no role in accounting for the decline in the inflow rate in this period.[26] Shift-share analyses—though informative—do not necessarily capture the full effects of demographic change. However, we find it useful because it helps us guide our examination of drivers of the decline in the inflow rate.

In light of our accounting exercise, and building on the earlier literature on female labor supply and firm dynamics, we now turn to the analysis of three channels that we show are important drivers of the downward trend in the inflow rate.[27]

V.A. The Increased Labor Force Attachment of Females

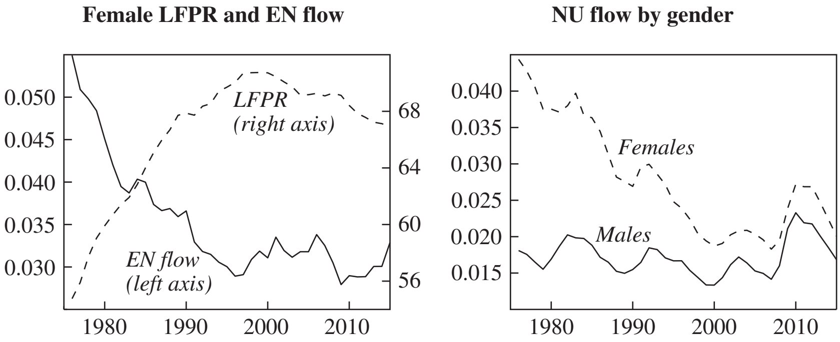
The United States experienced *grand gender convergence* in the 20th century, with female labor participation increasing from about 47 percent in 1976 to about 60 percent in 2000 (Goldin 2006).[28] The main driver of the rise in the female participation rate was the increase in participation of married females with children. Females started to work longer into their pregnancy, and they started working after childbirth sooner than their counterparts in the 1960s, likely due to changes in social norms, more widespread availability of maternity leave, and advances in maternal health and child care. As labor market interruptions related to childbearing declined, females' labor force attachment gradually

26. This observation is consistent with the findings of Davis and Haltiwanger (2014), who show using various data sources that similar patterns apply to broader measures of worker allocation such as hires and separations.

27. It is possible that secular changes in factors affecting matching efficiency could have also played a role; however, these factors primarily affect the outflow rate which exhibits only a mild secular trend.

28. As discussed extensively by Goldin (2006, and the references therein), a large body of literature examines the drivers of the dramatic rise in female labor force participation. Various drivers of this change have been identified such as technological change, contraceptive innovation (for example, the birth control pill), a shift to a unilateral divorce legal framework, changes in social attitudes and norms toward married females working, advances in home production technology, a decline in maternal mortality, and the introduction of infant formula.

Figure 13. Labor Force Participation and Flows, 1976–2015[a]



Source: Barnichon and Mesters (2018).

a. LFPR = labor force participation rate; EN = employment-to-nonparticipation; NU = nonparticipation-to-unemployment. Female labor force participation rate and EN flow rate (left) and NU flow rate normalized by the labor force by gender (right) for age 16–65 years.

increased.[29] The left panel of figure 13 shows that employed females left the labor force at a much lower rate in the 1990s than in the late 1970s. Though we do not have labor market flows before 1976, tabulations by Stephen Marston (1976) showed that the employment-to-nonparticipation flow rate for white females age 25–59 was 4.76 percent, while it was only 0.37 percent for white males in the same age category for the period 1967–73. Marston (1976) argues that the high rate at which employed females leave the labor force was the main factor in the higher unemployment rates they experienced. Marston referred to this as possibly a consequence of *participation instability*, almost antonymous to *increased labor force attachment*.

This decline in labor force exits has important implications for unemployment. Having uninterrupted employment spells allows workers to build more stable employment relationships, which is likely to reduce frictional unemployment through a decline in the incidence of job loss and the incidence of unemployment during reentry into the labor force. Examination of gross-flows data from the CPS based on longitudinally matched monthly CPS micro data confirms this intuition. As the right panel of figure 13 shows, as labor force departures became less common for females, entry from being out of the labor force into unemployment also became

29. For detailed statistics characterizing the changing patterns, see U.S. Census Bureau (2008).

increasingly rare. In the late 1970s, unemployment inflows from nonparticipation were about 4.5 percent of the labor force and declined by more than half, to about 2 percent by the late 1990s.[30]

Females also became less likely to leave unemployment for nonparticipation, increasing their duration of unemployment. Consequently, both the inflow and outflow gaps disappeared. On net, the decline in unemployment inflows dominated the rise in the duration of unemployment, causing full convergence of the unemployment rate of females to levels similar to males.[31] To summarize, even though declining exits from employment to nonparticipation will not have an immediate effect on the unemployment rate, they affected females' unemployment rate by lowering frictional unemployment, as shown by Katharine Abraham and Shimer (2002) using a flow decomposition and by Albanesi and Şahin (2018) using a three-state search and matching model.

GRAND GENDER CONVERGENCE IN THE CROSS-STATE DATA Evidence from U.S. states also confirms the relationship in the aggregate data: the rise in the female labor force participation rate was accompanied by an increase in labor force attachment, which in turn reduced frictional unemployment for females, generating a full convergence of unemployment rates by gender. We examine the evolution of the gender gaps in unemployment inflows and outflows at the state level. We first define the gender participation rate gap for each state at time t as

$$\frac{lfpr_{t,m} - lfpr_{t,f}}{lfpr_{t,m}}$$

and the unemployment inflow and outflow gaps as

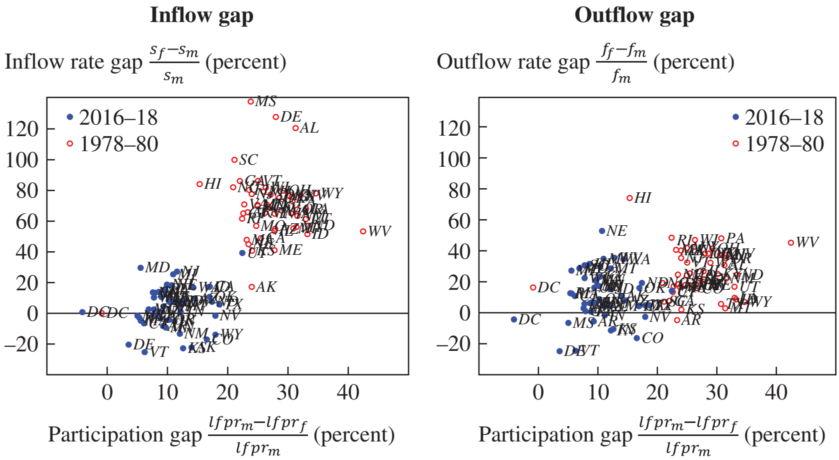
$$\frac{s_{t,f} - s_{t,m}}{s_{t,m}} \quad \text{and} \quad \frac{f_{t,f} - f_{t,m}}{f_{t,m}},$$

with m denoting male outcomes and f denoting female workers' outcomes.

30. Abraham and Shimer (2002) and Albanesi and Şahin (2018) show nonparticipation to unemployment flows as a fraction of the stock of nonparticipation. We normalize these flows by the labor force because the unemployment rate is measured as a share of the labor force. An alternative is to compute the unemployment inflows by reason of unemployment, following Elsby, Michaels, and Solon (2009), who found that the inflow rate for labor force entrants declined starting in the early 1980s.

31. Albanesi and Şahin (2018) show that, though the gender unemployment gap has disappeared, the relative cyclicality of unemployment by gender has not changed.

Figure 14. The Convergence of the Inflow and Outflow Gaps



Source: Current Population Survey.
a. Participation and unemployment inflow/outflow gaps are for the 1978–80 and 2016–18 periods.

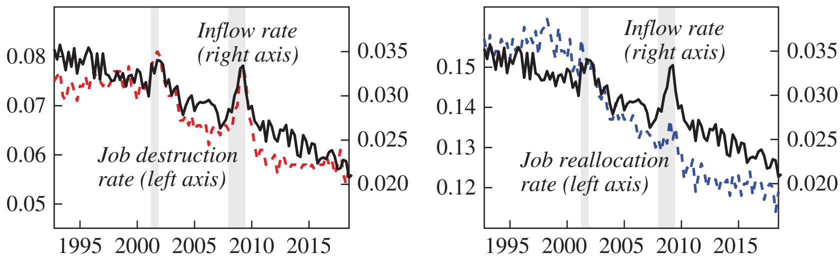
Figure 14 shows the state-level participation gaps and unemployment inflow and outflow gaps for the 1978–80 and 2016–18 periods. The convergence in labor market outcomes is clear. Moreover, unemployment inflows exhibit a starker convergence over time, consistent with the patterns in the aggregate data.

V.B. The Decline in Job Destruction Reallocation

Although females play an important role in the evolution of unemployment flows, almost all demographic groups' inflows declined over time. Especially after 1996, declines in group-specific inflow rates were the sole driver of the decline in the inflow rate, suggesting a common factor. Moreover, the rate employed workers transitioned into unemployment declined for both males and females, despite the dramatic job destruction at the onset of the Great Recession. This pattern suggests that changes in labor demand factors likely played a role.

The decline in unemployment inflows coincided with the decline in the volatility of firm-level growth rates and job destruction, as shown by Davis and others (2006); see figure 15. Search and matching models provide a natural link between the intensity of shocks that firms face and the incidence of unemployment. In this class of models with an endogenous job destruction margin, a decline in the intensity of firm-level

Figure 15. Job Destruction/Reallocation and Unemployment Inflow Rate, 1995–2018[a]



Sources: Current Population Survey; Business Employment Dynamics.
a. This figure shows job destruction and inflow rates (left) and job reallocation and inflow rates (right).

shocks would lower job destruction and the incidence of unemployment (Mortensen and Pissarides 1994). Davis and others (2010) formally examined this hypothesis, showing that industry-level movements in unemployment inflows are closely related to industry-level movements in several indicators for the intensity of idiosyncratic shocks for the 1990–2004 period. In this subsection, we extend Davis and others' (2010) analysis to the 1991–2017 period and evaluate the role of declining volatility on the trend decline in unemployment inflows and the employment-to-unemployment transition rate.

We use the Business Employment Dynamics (known as BED) data, which provide quarterly measures of job destruction at the industry level. We follow Davis and others (2010) and aggregate the data to these broad industry groups: construction, manufacturing, transportation and utilities, retail and wholesale trade, FIRE (finance, insurance, and real estate), and services; and we exploit within-industry time variation, the preferred specification of Davis and others (2010). The job destruction rate from quarter $t - 1$ to t is computed as the sum of job losses that are the result of contractions in employment at existing establishments and the loss of jobs at closing establishments, and it is expressed as a rate by dividing through by total employment.

We find economically and statistically significant effects of job destruction and job reallocation on the inflow rate and the employment-to-unemployment transition rate (table 3). The decline in job destruction and reallocation could be interpreted as declining firm level volatility and could arise from a changing nature of shocks or the declining responsiveness to shocks by firms as in Faberman (2017) and Ryan Decker and others (2017).

Table 3. Unemployment Inflow Rate and Employment-to-Unemployment Transition Rate Regressed on Job Destruction and Job Reallocation Rates, Quarterly Data[a]

	Inflow rate		*Employment-to-unemployment flow rate*	
Regressor	*(1)*	*(2)*	*(1)*	*(2)*
Job destruction rate	0.448***		0.382***	
	(0.0266)		(0.0202)	
Job reallocation rate		0.240***		0.197***
		(0.0198)		(0.0128)
Observations	618	618	618	618
R^2	0.935	0.927	0.950	0.933

Sources: Current Population Survey; Business Employment Dynamics.

a. Robust standard errors are in parentheses; ***$p < 0.01$, **$p < 0.05$, *$p < 0$. Quarterly data are from 1992:Q3–2018:Q1. The table includes time and sector fixed effects, with seven industry sectors.

V.C. Dual Aging

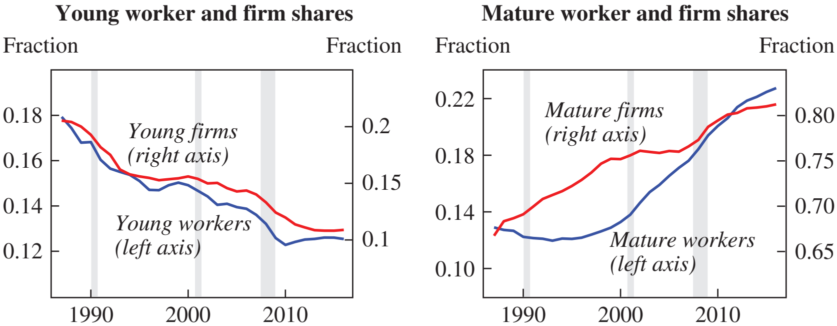
The U.S. economy has been experiencing a striking shift toward older workers and older firms since the mid-1990s, as we have discussed in section II. About 18 percent of the labor force was made up of workers between 16 to 24 years (*young workers*) in 1987. By 2017, this fraction declined to about 12 percent.[32] *Young firms'* (firms younger than 5 years old) employment share also followed a similar pattern, with their employment share declining from about 20 percent to 10 percent. On the flip side, in 1987, firms 11 or more years old—which we call *mature firms*, following Benjamin Pugsley and Şahin (2019)—used to employ about two-thirds of the workers in the economy. By 2017, this fraction increased to about 80 percent, as seen in figure 16.

Both worker age and firm age are widely recognized as important observables in accounting for differences in economic outcomes of workers and firms.[33] Table 4 shows the average unemployment inflow rates by worker age and job destruction rates by firm age. Younger workers are more than four times more likely to flow into unemployment than prime age workers. Similarly, firms between one and five years old are almost twice as likely to destroy jobs as their older counterparts. These patterns

32. In addition to the aging of the labor force, the ongoing decline in young workers' participation rate was a factor in this notable decline. As Krueger (2017) argues, the decline in participation of young workers was mostly offset by an increase in their college enrollment.

33. For worker age, see Perry (1970) and Shimer (1998); and for firm age, see Haltiwanger, Jarmin, and Miranda (2013) and Fort and others (2013).

Figure 16. The Dual Aging of Workers and Firms, 1987–2016[a]



Sources: Current Population Survey; Business Dynamics Statistics.
a. This figure shows the employment share of 16- to 24-year-old workers and the employment share of 5-year-old and younger firms and employment shares of 55-and-above workers and employment share of firms 11 or more years old.

Table 4. Inflow and Job Destruction Rates, by Age[a]

Monthly inflow rates by worker age, 1978–2018		*Annual job destruction rates by firm age, 1987–2016*	
16–24	0.097	1–5	0.244
25–54	0.023	6–10	0.176
55+	0.015	11+	0.124

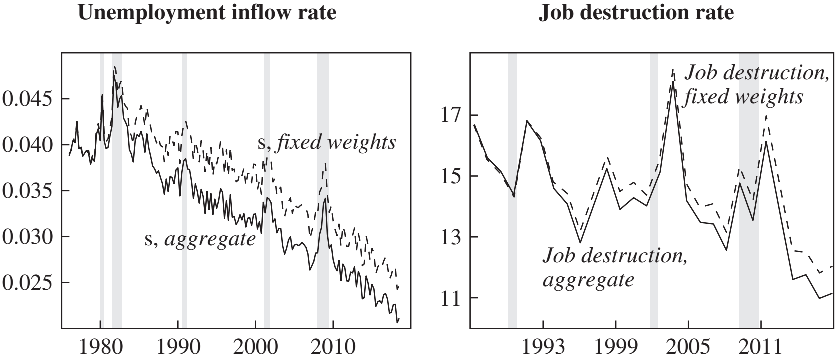
Sources: Current Population Survey (CPS); Business Dynamics Statistics (BDS).
a. This figure shows the average unemployment inflow rate by worker age in the CPS for 1976–2018 and the average job destruction rate by firm age in the BDS for 1987–2016. The inflow rates are calculated using the CPS, and the job destruction rates are calculated using the BDS.

suggest that a direct consequence of dual aging is a decline in unemployment inflows and job destruction.

We first conduct a simple worker-age-composition adjustment in the left panel of figure 17. We set the age composition of workers to their 1976 shares. We use three age groups for workers: 16–24, 25–54, and 55 or older. The shift toward an older population by itself accounts for about a quarter of the decline in the inflow rate, yet attributes a significant portion to the age-specific evolution of the inflow rate—a finding that resonates with our earlier analysis.

We repeat the same simple firm-age-composition adjustment in the right panel of figure 17, setting the age composition of firms to their 1987 shares using the Business Dynamics Statistics data set. We use three age groups of firms: 1–5 years, 6–10 years, and 11 or more years. Though it is hard

Figure 17. Shift Share Analyses with Fixed Age Composition, Various Periods[a]



Sources: Current Population Survey; Business Dynamics Statistics; authors' calculations.
a. Unemployment inflow rate: aggregate and keeping the worker age composition unchanged at its 1976 shares (left) and job destruction rate: aggregate and keeping the firm age composition unchanged at its 1987 shares (right).

to assess the exact fraction that this calculation accounts for due to the pronounced countercyclicality of the job destruction rate, the change in the firm age compositions seems to be about as important as worker aging. However, the bulk of the decline still remains unaccounted for, similar to the inflow rate. This finding is consistent with those of Davis and Haltiwanger (2014) and of Haltiwanger, Ron Jarmin, and Javier Miranda (2013), who also show that though the shifts in the worker and firm age compositions go in the right direction, they still remain short of explaining the majority of the decline in the unemployment inflow and job destruction rates.

Although the shift in worker and firm age composition falls short of accounting for the decline in the inflow rate, recent research has emphasized that aging could also influence the economy by affecting age-specific outcomes, as noted by Shimer (2001), Fatih Karahan and Serena Rhee (2017), and Niklas Engbom (2019). Shimer (2001) refers to the direct effect as the effect of aging arising solely from changes in the age composition and to any additional effects as the indirect effect. These papers argue that the effect of aging goes beyond just shifting the composition of the economy in the context of unemployment, migration, and various measures of dynamism.

We build on this insight and show that the age composition of workers affects age-specific inflow rates and that the age composition of firms affects firm-age-specific job destruction rates, suggesting that indirect effects also play a substantial role.

DUAL AGING IN THE CROSS-STATE DATA We now turn to geographic variations to examine the direct and indirect effects of dual aging on unemployment inflows and job destruction rates using cross-state data.

Worker demographics and unemployment inflows. We should expect those states with larger changes in demographic makeup to experience the sharpest declines in inflow rates. Given the slow-moving nature of demographic changes, it is natural to compare long-horizon changes in these variables. To do so, we regress the change in the inflow rate for each state from its average value in 1978–82 to its average value in 1997–2001 on the change in the share of those age 15–24 relative to those age 15–64 from 1978 to 1998. We choose to take five-year averages of the inflow rates, and these years in particular, to ensure that our long differences are not unduly affected by the state of the business cycle (the first year in our sample is 1978, and the subsequent recession began in January 1980). We focus on the period up until the late 1990s because that is the period during which the share of young people in the population moved dramatically; since then, the changes have been relatively modest. In table 5, we show the results for this long-difference regression:

$$(41) \qquad s^i_{1997-2001} - s^i_{1978-1982} = \beta_0 + \beta_1 \left[\left(\frac{\text{pop 15 to 24}}{\text{pop 15 to 64}} \right)_{i,1998} - \left(\frac{\text{pop 15 to 24}}{\text{pop 15 to 64}} \right)_{i,1978} \right] + \varepsilon_i .$$

The ordinary least squares (OLS) estimate suggests that a decline of 1 percentage point in the share of young people in a state corresponds to a fall of about 0.15 percentage point in the inflow rate. A common choice of instrument in regressions with age shares is to use lagged age shares adjusted for the deterministic aging that would be expected to occur.[34] This strategy relies on the idea that lagged age shares are not informative about current business conditions—such as a labor demand shock—that could potentially move both the age composition and unemployment inflows contemporaneously. In this case, we need to forecast the share of young people in 1998 as of 1978. We use 1978 births along with the rest of the age distribution at the time to do so. In particular, we forecast the long-horizon change by replacing

34. In a recent example, Davis and Haltiwanger (2014), estimate the effect of reallocation measures on employment and unemployment outcomes by age, gender, and education, using instruments based on age shares at the state level.

Table 5. Changes in Inflow Rates and Population Composition[a]

	Long-horizon change in inflow rate for:				
Specification	*Overall*	*All ages*	*16–24*	*25–54*	*55+*
Ordinary least squares	0.157	0.089	0.143	0.076	0.049
p value	(0.006)	(0.260)	(0.489)	(0.051)	(0.641)
IV	0.186	0.174	0.334	0.099	0.087
90% confidence interval	[0.05, 0.33]	[–0.00, 0.36]	[–0.17, 0.88]	[0.01, 0.20]	[–0.06, 0.24]
Observations	50	148	50	50	48

Sources: Current Population Survey; National Cancer Institute.

a. IV = instrumental variable. This table reports regression results for the specification of equation 41. The second row reports p values associated with the ordinary-least-squares estimate with robust standard errors; the fourth row reports weak-instrument robust confidence intervals constructed by inverting the Anderson–Rubin test (Mikusheva and Poi 2006). The “all ages” specification includes age effects. Results omit the District of Columbia.

$$\left(\frac{\text{pop 15 to 24}}{\text{pop 15 to 64}}\right)_{i,1998} - \left(\frac{\text{pop 15 to 24}}{\text{pop 15 to 64}}\right)_{i,1978}$$

with

$$\left(\frac{6 \cdot (\text{1978 births}) + \text{pop 1 to 4}}{6 \cdot (\text{1978 births}) + \text{pop 1 to 44}}\right)_{i,1978} - \left(\frac{\text{pop 15 to 24}}{\text{pop 15 to 64}}\right)_{i,1978}.$$

In words, to estimate the population of 15- to 24-year-olds in 1998 requires an estimate of the number of births in the period 1979–83 along with births and the population of 1- to 4-year-olds in 1978. We estimate the number of births in 1979–83 by assuming that births are constant at their 1978 level over that period. The instrumental variables (IV) estimate, also shown in table 5, is slightly larger than the OLS estimate, and weak-IV robust confidence intervals comfortably reject the null of a slope of zero.[35] These preceding results are similar in spirit to that of the shift-share analysis we have already conducted in table 2.

We next replace the long-horizon differences in inflow rates for all workers by the corresponding changes for workers below 25, age 25–54, and those 55 and over. This allows us to assess whether the maturing population—the decline in young people—is correlated with declines in

35. Although we employ weak-instrument robust confidence intervals, we note for reference that the first-stage F statistic is about 90.

the inflow rate for prime age and older workers. The results are suggestive that the change in the share of young people is associated with changes in separation rates across age groups (the second column of table 5). If we dig deeper and run these regressions by each age group separately, we observe that a higher share of young people is positively associated to the inflow rates of all groups with the effect declining by age (last three columns of the table). The most striking finding perhaps is the observation that prime age workers' unemployment inflow rate declines with the share of young workers in the economy. These results suggest that separation rates of prime age workers, in particular, were affected by the maturing population.

Firm demographics and job destruction. To investigate the role of firm aging, we follow a similar empirical strategy as in the case of worker demographics. We should expect those states with more substantial shifts toward older firms to experience the biggest declines in job destruction. We again consider long changes in job destruction due to the slow-moving firm demographics, as shown in figure 16. We compare the 3-year average of job destruction rates in 1987–89 with 2012–14 and examine how they are affected by the aging of firms, using the change in employment share of firms 11 years and older as a proxy for aging. This choice of regressor is motivated by the work of Haltiwanger, Jarmin, and Miranda (2013), who show that most of the young-firm dynamics continue throughout the first 10 years of firms' lives.

In table 6, we show the results for this long-difference regression, using OLS:

$$(42)\quad jd^{i}_{2012-2014} - jd^{i}_{1987-1989} = \beta_0 + \beta_1\left[\left(\frac{\text{emp}\,11+}{\text{emp}}\right)_{i,2014} - \left(\frac{\text{emp}\,11+}{\text{emp}}\right)_{i,1987}\right] + \varepsilon_i,$$

using job destruction data by state from Business Dynamics Statistics. The OLS estimate implies that an increase of 1 percentage point in the employment share of mature firms in a state corresponds to a fall of about 0.28 percentage point in the job destruction rate. This effect is both statistically and economically significant, and is more substantial quantitatively than the implication of the shift-share analysis. However, it is subject to the usual critique that firm demographics and job destruction could be affected by common shocks. To address this concern, we devise an IV strategy that parallels the IV approach that we have employed for worker demographics. To do so, we use the employment share of new firms (births) in

Table 6. Changes in Job Destruction Rates and Firm Age Composition[a]

	Long-horizon change in job destruction rate for:				
Specification	*Overall*	*All ages*	*1Y–5Y*	*6Y–10Y*	*11Y+*
Ordinary least squares	–0.284	–0.254	–0.261	–0.303	–0.197
p value	(0.008)	(0.002)	(0.146)	(0.025)	(0.035)
IV	–0.492	–0.571	–0.805	–0.595	–0.312
90% confidence interval	[–0.73, –0.33]	[–0.76, –0.41]	[–1.35, –0.45]	[–0.95, –0.34]	[–0.50, –0.16]
Observations	50	150	50	50	50

Source: Business Dynamics Statistics.

a. IV = instrumental variable. This table reports regressions results for the specification of equation 42. The second row reports *p* values associated with the ordinary-least-squares estimate with robust standard errors; the fourth row reports weak-instrument robust confidence intervals constructed by inverting the Anderson–Rubin test (Mikusheva and Poi 2006). The "all ages" specification includes age effects. Results omit the District of Columbia.

1979 as the instrument.[36] Petr Sedlácek and Vincent Sterk (2017) illustrate the strong persistence over time in the employment shares of startups. Because surviving firms that originated in 1979 did not become age 11 years until after 1987, these lagged employment shares of startups should forecast the subsequent long-horizon change of employment shares in old firms.[37] The bottom rows of table 6 show that the IV estimate is even stronger, suggesting a more substantial effect of firm aging on job destruction. In fact, the OLS estimate is outside the 90 percent weak-instrument robust confidence interval.

We next assess the indirect effect of firm aging on job destruction by investigating the relation between firm demographics and job destruction rates by firm age. In the second column of table 6, we pool the job destruction rates for all three age groups and include age effects. In concert with the overall results, the IV estimates suggest a larger magnitude of effect. In the table's last three columns, we consider individual specifications for each age group. We see clear negative effects of firm aging on job destruction for all firm age groups, suggesting that the shift in the age composition of firms is not the only effect of aging on job destruction. An older firm age distribution implies a lower overall job destruction rate by lowering job destruction for firms of all ages. We should also note that, across five specifications, the p values associated with the IV estimates are all smaller than 0.01. To ensure that our results are not driven by, for example, the types of industries that prevail in each state, we also report results for a panel version of equation 42, splitting the long-horizon change into observations of changes over two subperiods and taking the change (in the change) to account for unobserved heterogeneity (see table C.3 in the online supplemental appendix). Although the confidence intervals widen considerably, we continue to observe an estimated negative relationship and reject the null hypothesis at least at the 10 percent level, as judged by the weak-IV robust confidence intervals for all but the 6Y–10Y age category.

Our analysis showed that changes in worker and firm demographics, to which we refer as the dual aging of the U.S. economy, are important drivers of the decline in job destruction and unemployment inflows—two measures that we linked in the preceding subsection. Although the change in worker demographics is directly attributable to the baby boom, the drastic increase in births after World War II, the emphasis on aging of

36. Although the Business Dynamics Statistics data begin in 1977, we use 1979, due to measurement concerns discussed by Pugsley and Şahin 2019.

37. Although we employ weak-instrument robust confidence intervals, we note for reference that the first-stage F statistic is about 28.

firms is relatively new because data have only recently become available. However, Pugsley and Şahin (2019), using a firm dynamics framework, showed that the intuition is very similar for firms: declining firm births almost fully account for the shift of employment toward older firms. Moreover, Karahan, Pugsley, and Şahin (2018) show that the origin of the decline in firm entry is the decline in the labor supply growth arising from the aging of the baby boom cohort and the flattening out of the female labor force participation rate. These downward trends in unemployment inflows and job destruction pertain to a broader set of worker and job reallocation measures, as first documented by Davis and others (2006). Relatedly, recent research by Davis and Haltiwanger (2014) has shown that, in addition to shifts to older businesses and an aging workforce, policy developments that suppress reallocation—such as occupational labor supply restrictions, exceptions to the employment-at-will doctrine, the establishment of protected worker classes, and job lock associated with employer-provided health insurance—are among the policy factors that suppress labor market fluidity. Although analyses of these factors are beyond the scope of our paper, we believe that the interaction of policy decisions with labor market reallocation is an important issue for better understanding many important aggregates, such as unemployment, employment, productivity, and wages.

VI. Conclusion

We estimate the natural rate of unemployment in the 1960–2018 period by unifying two distinct estimation approaches that are popular in the literature. We exploit a rich set of labor market and inflation expectations data to provide tight estimates of the natural rate and study the underlying determinants of its movements. As of the third quarter of 2018, we estimate that u_t^* was about 4 percent; in particular, using only information from price inflation, we estimate that u_t^* stood at 4.0 percent with a 68 percent confidence interval of 3.5 to 4.5 percent. When we add information from wage inflation, the estimate shifts down slightly, to 3.8 percent, with an associated confidence interval of 3.5 to 4.2 percent. Our natural rate estimate is about 60 basis points lower than that of the CBO's estimate and 50 basis points lower than the median longer-run unemployment rate projection from the Federal Open Market Committee's "Summary of Economic Projections."[38] Importantly, our estimates imply that the unemployment gap was roughly closed toward the end of 2018.

38. See CBO (2019); FOMC (2019).

During the Great Recession, we find that the unemployment gap peaked at about 4 percentage points, which was far more severe than in any other downturn since the 1960s. Moreover, the closing of the unemployment gap has occurred only slowly, falling below 2 percentage points in 2014—about five years after the recession ended, and closing entirely only recently. We confront the micro data–based estimates of the rise in the natural rate during the Great Recession and find that our estimate of the rise is, remarkably, in agreement with the rise in mismatch unemployment and the decline in recruiting intensity. We view this similitude as an important success, given that these measures use almost completely separate sources of information.

Our analysis highlights a slow-moving secular trend that has been dragging down the unemployment rate since the early 1980s. This downward trend, until the late 1990s, was mostly driven by young workers and prime age females, while the secular trend in the last two decades is common across age and gender groups. We identify the rise in female labor force attachment, the decline in job destruction and reallocation intensity, and the dual aging of workers and firms in the economy as key drivers of this trend. Furthermore, we view these three developments as major changes that have had, and will continue to have, important and long-lasting effects on the economy.

The female labor force participation rate flattened in the late 1990s, and the unemployment rate for females fully converged with that of males. The participation gap has improved minimally since then, mostly on account of the deterioration of male participation outcomes and a gap of about 14 percentage points that still exists between prime age males and females. Our analysis of labor force attachment suggests that declining male attachment will be an upside risk for unemployment in the future, even though its effects are, thus far, more than being offset by the downward trend in job destruction. Another implication of our findings is that improvements in child care availability and maternity leave policy for females would also lower the natural rate of unemployment to the extent that they increase females' labor force attachment.

The aging of the population was predictable as early as the 1960s, and its consequences for innovation, productivity, government budgets, tax policy, Social Security, the labor market, and political economy have inspired an abundance of analyses and policy recommendations.[39] Although the discussion of the effects of aging goes back decades, there is still room

39. An insightful article by Cutler and others (1990) lays out various issues related to aging.

for further research on this topic, given that its effects on other parts of the economy—such as the decline in firm entry and the aging of firms—have taken shape. Another important implication of aging is the decline in workers' bargaining power, as recently analyzed by Andrew Glover and Jacob Short (2018). Using cross-sectoral variation, they find that older workers receive a smaller share of their marginal product than do younger workers, and they link the recent demographic trends in the United States to the declining bargaining power of workers. Because both worker and firm demographics are slow moving, and would likely take a long time to reverse, we expect these effects to persist.

Admittedly (and hopefully), our paper is not the last word on the natural rate of unemployment. However, we view our unified framework as a useful tool for future policy analyses, because it provides a bridge between the Phillips curve literature and the macro-labor literature, which focuses on measuring labor market efficiency by exploiting rich cross-sectional information. Moreover, the development of detailed micro data sources is a relatively recent development, and we expect that further progress harmonizing these two approaches will be made in future research.

ACKNOWLEDGMENTS The authors thank the *Brookings Papers* editors, Janice Eberly and James Stock, and commenters Steven Davis and Giorgio Primiceri. They also thank Olivier Blanchard, Jason Faberman, Bart Hobjin, Marco Del Negro, Domenico Giannone, Nora Traum, John Williams, and the Brookings conference participants for helpful comments. Réne Chalom, Brendan Moore, and Jin Yan provided excellent research assistance.

References

Abraham, Katharine G., and Robert Shimer. 2002. "Changes in Unemployment Duration and Labor Force Attachment." In *The Roaring Nineties*, edited by Alan Krueger and Robert Solow. New York: Russell Sage Foundation.

Abrahams, Michael, Tobias Adrian, Richard K. Crump, Emanuel Moench, and Rui Yu. 2016. "Decomposing Real and Nominal Yield Curves." *Journal of Monetary Economics* 84: 182–200.

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

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

Ball, Laurence, and Sandeep Mazumder. 2011. "The Evolution of Inflation Dynamics and the Great Recession." *Brookings Papers on Economic Activity*, Spring, 337–405.

Barnichon, Regis, and Geert Mesters. 2018. "On the Demographic Adjustment of Unemployment." *Review of Economics and Statistics* 100, no. 2: 219–31.

Barnichon, Regis, and Christopher J. Nekarda. 2012. "The Ins and Outs of Forecasting Unemployment: Using Labor Force Flows to Forecast the Labor Market." *Brookings Papers on Economic Activity*, Fall, 83–131.

Blanchard, Olivier. 2016. *The U.S. Phillips Curve: Back to the '60s?* Policy Brief PB16-1. Washington: Peterson Institute for International Economics. https://piie.com/publications/policy-briefs/us-phillips-curve-back-60s.

Blanchard, Olivier J., and Peter Diamond. 1989. "The Beveridge Curve." *Brookings Papers on Economic Activity*, no. 1: 1–76.

———. 1990. "The Cyclical Behavior of the Gross Flows of U.S. Workers." *Brookings Papers on Economic Activity*, no. 2: 85–155.

Blanchard, Olivier J., and Lawrence H. Summers. 1986. "Hysteresis and the European Unemployment Problem." In *NBER Macroeconomics Annual*, edited by Stanley Fischer. Cambridge, Mass.: MIT Press.

BLS (Bureau of Labor Statistics). 2018. "Labor Force Statistics from the Current Population Survey." https://www.bls.gov/cps/definitions.htm.

Brauer, David. 2007. "The Natural Rate of Unemployment." Working Paper 2007-06. Washington: Congressional Budget Office. https://www.cbo.gov/publication/18566.

Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12, no. 3: 383–98.

Carter, C. K., and R. Kohn. 1994. "On Gibbs Sampling for State Space Models." *Biometrika* 81, no. 3: 541–53.

Carvalho, Carlos, Sefano Eusepi, Emanuel Moench, and Bruce Preston. 2017. "Anchored Inflation Expectations." Working paper. https://dx.doi.org/10.2139/ssrn.3018198.

CBO (Congressional Budget Office). 2019. "Chapter 2: The Economic Outlook." In *The Budget and Economic Outlook: 2019 to 2029*. Washington: CBO. https://www.cbo.gov/system/files?file=2019-01/54918-Outlook-Chapter2.pdf.

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.

Cutler, David M., James M. Poterba, Louise M. Sheiner, and Lawrence H. Summers. 1990. "An Aging Society: Opportunity or Challenge?" *Brookings Papers on Economic Activity*, no. 1: 1–73.

Daly, Mary C., Bart Hobijn, Ayşegül Şahin, and Robert G. Valletta. 2012. "A Search and Matching Approach to Labor Markets: Did the Natural Rate of Unemployment Rise?" *Journal of Economic Perspectives* 26, no. 3: 3–26.

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger. 2013. "The Establishment-Level Behavior of Vacancies and Hiring." *Quarterly Journal of Economics* 26, no. 2: 581–622.

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger, Ron Jarmin, and Javier Miranda. 2010. "Business Volatility, Job Destruction, and Unemployment." *American Economic Journal: Macroeconomics* 2, no. 2: 259–87.

Davis, Steven J., and John C. Haltiwanger. 2014. "Labor Market Fluidity and Economic Performance." NBER Working Paper 20479. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w20479.pdf.

Davis, Steven J., John Haltiwanger, Ron Jarmin, and Javier Miranda. 2006. "Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms." In *NBER Macro Economics Annual*, edited by Daron Acemoglu, Kenneth Rogoff, and Michael Woodford. Cambridge, Mass.: MIT Press.

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

Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2018. "Changing Business Dynamism and Productivity: Shocks vs. Responsiveness." Working Paper 2018-007. Washington: Federal Reserve Board of Governors. https://www.federalreserve.gov/econres/feds/files/2018007pap.pdf.

Del Negro, Marco, Marc P. Giannoni, Domenico Giannone, and Andrea Tambalotti. 2017. "Safety, Liquidity, and the Natural Rate of Interest." *Brookings Papers on Economic Activity*, Spring, 235–94.

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: 168–196.

Diamond, Peter A. "Mobility Costs, Frictional Unemployment, and Efficiency." *Journal of Political Economy* 89, no. 4 (1981): 798–812.

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin. 2010. "The Labor Market in the Great Recession." *Brookings Papers on Economic Activity*, Spring, 1–69.

———. 2013. "Unemployment Dynamics in the OECD." *Review of Economics and Statistics* 95, no. 2: 530–548.

Elsby, Michael W. L., Ryan Michaels, and Gary Solon. 2009. "The Ins and Outs of Cyclical Unemployment." *American Economic Journal: Macroeconomics* 1, no. 1: 84–110.

Engbom, Niklas. 2019. “Firm and Worker Dynamics in an Aging Labor Market.” Working Paper 756. Minneapolis: Federal Reserve Bank of Minneapolis. https://www.minneapolisfed.org/research/wp/wp756.pdf.

Erceg, Christopher J., Dale W. Henderson, Andrew T. Levin, 2000. “Optimal Monetary Policy with Staggered Wage and Price Contracts.” *Journal of Monetary Economics* 46, no. 2: 281–313.

Faberman, R. Jason. 2017. “Job Flows, Jobless Recoveries, and the Great Moderation.” *Journal of Economic Dynamics & Control* 76: 152–70.

Faberman, R. Jason, and Ayşegül Şahin. 2011. “Reserve Bank Report on Structural Unemployment.” FOMC Memo, Meeting of January 25–26. https://fraser.stlouisfed.org/title/677/item/22545/content/pdf/fomc20110101memo01/toc/521590.

FOMC (Federal Open Market Committee). 2011a. “Material for FOMC Briefing on Structural Unemployment.” January 25–26. https://www.federalreserve.gov/monetarypolicy/files/FOMC20110126material.pdf

———. 2011b. “Reserve Bank Report on Structural Unemployment.” https://www.federalreserve.gov/monetarypolicy/files/FOMC20110101memo01.pdf.

———. 2012. “Domestic Economic Developments and Outlook.” https://www.federalreserve.gov/monetarypolicy/files/FOMC20121212tealbooka20121205.pdf.

———. 2019. “Economic Projections of Federal Reserve Board Members and Federal Reserve Bank Presidents under Their Individual Assessments of Projected Appropriate Monetary Policy, September 2018.” September 26. https://www.federalreserve.gov/monetarypolicy/files/fomcprojtabl20180926.pdf.

Fort, Teresa C., John Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2013. “How Firms Respond to Business Cycles: The Role of Firm Age and Firm Size.” *IMF Economic Review* 61, no. 3: 520–59.

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

Fuhrer, Jeff, and George Moore. 1995. “Inflation Persistence.” *Quarterly Journal of Economics* 110, no. 1: 127–59.

Fujita, Shigeru, and Garey Ramey. 2009. “The Cyclicality of Separation and Job Finding Rates.” *International Economic Review* 50, no. 2: 415–30.

Galí, Jordi, 2011. “The Return of the Wage Phillips Curve.” *Journal of the European Economic Association* 9, no. 3: 436–61.

———. 2015. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*. Princeton, N.J.: Princeton University Press.

Galí, Jordi, and Mark Gertler. 1999. “Inflation Dynamics: A Structural Econometric Analysis.” *Journal of Monetary Economics* 44, no. 2: 195–222.

Galí, Jordi, Frank Smets, and Rafael Wouters. 2012. “Unemployment in an Estimated New Keynesian Model.” In *NBER Macroeconomics Annual*, edited by Daron Acemoglu and Michael Woodford. Cambridge, Mass.: MIT Press.

Gertler, Mark, Luca Sala, and Antonella Trigari. 2008. “An Estimated Monetary DSGE Model with Unemployment and Staggered Nominal Wage Bargaining.” *Journal of Money, Credit and Banking* 40, no. 8: 1713–64.

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.

Goldin, Claudia. 2006. "The Quiet Revolution That Transformed Women's Employment, Education, and Family." *American Economic Review* 96, no. 2: 1–21.

Goodfriend, Marvin, and Robert G. King. 1997. "The New Neoclassical Synthesis and the Role of Monetary Policy." In *NBER Macroeconomics Annual*, edited by Ben S. Bernanke and Julio Rotemberg. Cambridge, Mass.: MIT Press.

Gordon, Robert J. 1970a. "Prices in 1970: The Horizontal Phillips Curve?" *Brookings Papers on Economic Activity*, no. 1: 8–47.

———. 1970b. "The Recent Acceleration of Inflation and Its Lessons for the Future." *Brookings Papers on Economic Activity*, no. 1: 8–47.

———. 1972. "Wage-Price Controls and the Shifting Phillips Curve." *Brookings Papers on Economic Activity*, no. 2: 385–421.

———. 1973. "The Response of Wages and Prices to the First Two Years of Controls." *Brookings Papers on Economic Activity*, no. 3: 765–79.

———. 1982. "Inflation, Flexible Exchange Rates, and the Natural Rate of Unemployment." In *Workers, Jobs, and Inflation*, edited by Martin N. Baily. Washington: Brookings.

Greenspan, Alan. 1998. "Question: Is There a New Economy?" Speech given at Haas Annual Business Faculty Research Dialogue, Berkeley, September 4. https://www.federalreserve.gov/boarddocs/speeches/1998/19980904.htm.

Hall, Robert E. 1970a. "The Process of Inflation in the Labor Market." *Brookings Papers on Economic Activity*, no. 2: 343–93.

———. 1970b. "Why Is the Unemployment Rate So High at Full Employment?" *Brookings Papers on Economic Activity*, no. 3: 369–410.

———. 2005. "Employment Fluctuations with Equilibrium Wage Stickiness." *American Economic Review* 95, no. 1: 50–65.

———. 2011. "The Long Slump." *American Economic Review* 101, no. 2: 431–69.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda. 2013. "Who Creates Jobs? Small versus Large versus Young." *Review of Economics and Statistics* 95, no. 2: 347–61.

Hamilton, James D. 2005. "What's Real About the Business Cycle?" *Federal Reserve Bank of St. Louis Review* 87, no. 4: 435–52.

Haubrich, Joseph, George Pennacchi, and Peter Ritchken. 2012. "Inflation Expectations, Real Rates, and Risk Premia: Evidence from Inflation Swaps." *Review of Financial Studies* 25, no. 5: 1588–1629.

Herbst, Edward P., and Frank Schorfheide. 2015. *Bayesian Estimation of DSGE Models*. Princeton, N.J.: Princeton University Press.

Hornstein, Andreas, and Marianna Kudlyak. 2019. "Aggregate Labor Force Participation and Unemployment and Demographic Trends." Working Paper 2019-07. San Francisco: Federal Reserve Bank of San Francisco. https://www.frbsf.org/economic-research/files/wp2019-07.pdf.

Humphrey, Thomas M. 1991. "Nonneutrality of Money in Classical Monetary Thought." *Federal Reserve Bank of Richmond Economic Review* 18: 3–15.

Karahan, Fatih, Benjamin Pugsley, and Ayşegül Şahin. 2018. "Demographic Origins of the Startup Deficit." Working paper. https://goo.gl/A3eX4r.

Karahan, Fatih, and Serena Rhee. 2017. "Population Aging, Migration Spillovers, and the Decline in Interstate Migration." Working Paper 699. New York: Federal Reserve Bank of New York. https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr699.pdf.

Kim, Chang-Jin, and Charles R. Nelson. 1999. *State-Space Models with Regime Switching*. Cambridge, Mass.: MIT Press.

Krueger, Alan B. 2017. "Where Have All the Workers Gone? An Inquiry into the Decline of the U.S. Labor Force Participation Rate." *Brookings Papers on Economic Activity*, Fall, 1–87.

Laubach, Thomas. 2001. "Measuring the NAIRU: Evidence from Seven Economies." *Review of Economics and Statistics* 83, no. 2: 218–31.

Lucas, Robert E., Jr. 1972. "Expectations and the Neutrality of Money." *Journal of Economic Theory* 4: 103–24.

Marston, Stephen T. 1976. "Employment Instability and High Unemployment Rates." *Brookings Papers on Economic Activity*, no. 1: 169–210.

Mikusheva, Anna, and Brian P. Poi. 2006. "Tests and Confidence Sets with Correct Size When Instruments Are Potentially Weak." *Stata Journal* 6, no. 3: 335–47.

Modigliani, Franco, and Lucas Papademos. 1975. "Targets for Monetary Policy in the Coming Year." *Brookings Papers on Economic Activity*, no. 1: 141–65.

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

Montgomery, Alan L., Victor Zarnowitz, Ruey S. Tsay, and George G. Tiao. 1998. "Forecasting the U.S. Unemployment Rate." *Journal of the American Statistical Association* 93, no. 422: 478–93.

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.

Orphanides, Athanasios, and John C. Williams. 2002. "Robust Monetary Policy Rules with Unknown Natural Rates." *Brookings Papers on Economic Activity*, no. 2: 63–145.

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

———. 1972. "Unemployment Flows in the U.S. Labor Market." *Brookings Papers on Economic Activity*, no. 2: 245–92.

———. 1978. "Slowing the Wage–Price Spiral: The Macroeconomic View." *Brookings Papers on Economic Activity*, no. 2: 259–99.

Phelps, Edmund S. 1967. "Phillips Curves, Expectations of Inflation, and Optimal Unemployment over Time." *Economica* 34, no. 135: 254–81.

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

Polivka, Anne E., and Stephen M. Miller. 1998. "The CPS after the Redesign: Refocusing the Economic Lens." In *Labor Statistics Measurement Issues*, edited by John Haltiwanger, Marilyn E. Manser, and Robert Topel. Cambridge, Mass.: National Bureau of Economic Research.

Pugsley, Benjamin W., and Ayşegül Şahin. 2019. "Grown-Up Business Cycles." *Review of Financial Studies* 32, no. 3: 1102–47.

Roberts, John M. 1995. "New Keynesian Economics and the Phillips Curve." *Journal of Money, Credit and Banking* 27, no. 4: 975–84.

Rogerson, Richard. 1997. "Theory Ahead of Language in the Economics of Unemployment." *Journal of Economic Perspectives* 11, no. 1: 73–92.

Rotemberg, Julio J. 1987. "The New Keynesian Microfoundations." In *NBER Macroeconomics Annual*, edited by Stanley Fischer. Cambridge, Mass.: MIT Press.

Rotemberg, Julio J., and Michael Woodford. 1997. "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy." In *NBER Macroeconomics Annual*, edited by Ben S. Bernake. Cambridge, Mass.: MIT Press.

Şahin, Ayşegül, and Christina Patterson. 2013. "The Bathtub Model of Unemployment: The Importance of Labor Market Flow Dynamics." Liberty Street Economics Blog, March. https://libertystreeteconomics.newyorkfed.org/2012/03/the-bathtub-model-of-unemployment-the-importance-of-labor-market-flow-dynamics.html.

Şahin, Ayşegül, Joseph Song, Giorgio Topa, and Giovanni L. Violante. 2014. "Mismatch Unemployment." *American Economic Review* 104, no. 11: 3529–64.

Sargent, Thomas J. 1971. "A Note on the 'Accelerationist' Controversy." *Journal of Money, Credit and Banking* 3, no. 3: 721–25.

Sbordone, Argia M. 2002. "Prices and Unit Labor Costs: A New Test of Price Stickiness." *Journal of Monetary Economics* 49, no. 2: 265–92.

Schultze, Charles L. 1971. "Has the Phillips Curve Shifted? Some Additional Evidence." *Brookings Papers on Economic Activity*, no. 2: 452–67.

Sedláček, Petr, and Vincent Sterk. 2017. "The Growth Potential of Startups over the business Cycle." *American Economic Review* 107, no. 10: 3182–3210.

Shimer, Robert. 1998. "Why Is the U.S. Unemployment Rate So Much Lower?" In *NBER Macroeconomics Annual*, edited by Ben S. Bernake and Julio J. Rotemberg. Cambridge, Mass.: MIT Press.

———. 2001. "The Impact of Young Workers on the Aggregate Labor Market." *Quarterly Journal of Economics* 116, no. 3: 969–1007.

———. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review* 95, no. 1: 25–49.

———. 2012. "Reassessing the Ins and Outs of Unemployment." *Review of Economic Dynamics* 15, no. 2: 127–48.

Solow, Robert M. 1964. *The Nature and Sources of Unemployment in the United States*. Stockholm: Almqvist & Wiksell.

Staiger, Douglas, James H. Stock, and Mark W. Watson, 1997. “How Precise Are Estimates of the Natural Rate of Unemployment?” In *Reducing Inflation: Motivation and Strategy*, edited by Christina D. Romer and David H. Romer. 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.

Tasci, Murat. 2014. “The Ins and Outs of Unemployment in the Long Run: Unemployment Flows and the Natural Rate.” Working Paper 1233. Istanbul: Koç University-Tüsiad. https://eaf.ku.edu.tr/sites/eaf.ku.edu.tr/files/erf_wp_1233.pdf.

Tobin, James. 1974. “Monetary Policy in 1974 and Beyond.” *Brookings Papers on Economic Activity*, no. 1: 219–32.

U.S. Census Bureau. 2008. “2008 Current Population Report on Maternity Leave and Employment Patterns of First-Time Mothers: 1961–2008.” https://www.census.gov/prod/2011pubs/p70-128.pdf.

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, N.J.: Princeton University Press.

Yun, Tack. 1996. “Nominal Price Rigidity, Money Supply Endogeneity, and Business Cycles.” *Journal of Monetary Economics* 37 , no. 2: 345–70.

Comments and Discussion

COMMENT BY

STEVEN J. DAVIS Richard Crump, Stefano Eusepi, Marc Giannoni, and Ayşegül Şahin make an excellent contribution to our understanding of unemployment and inflation dynamics, in two main parts: First, they estimate a trend decline in the U.S. unemployment rate of about 3 percentage points since the early 1980s. This development reflects a secular fall in the unemployment inflow rate, which the authors link to several deeper forces. This part of the paper brings together and extends a wide range of previous studies. Second, they develop an empirical model around a Phillips Curve with forward- and backward-looking elements, their measured unemployment trend, and survey data on inflation expectations. They use their empirical model to estimate the natural rate of unemployment, u^*, and to interpret the joint evolution of labor market slack and inflation. In their characterization, u^* moves over time due to the evolution of the unemployment trend and due to temporary forces.

Here, I first elaborate on the deeper forces behind the downward drift in unemployment and explain why this drift is important, quite apart from its implications for inflation. Then I express doubts about the practical usefulness of Crump and colleagues' Phillips Curve model, and the Phillips Curve concept more broadly, for the conduct of monetary policy. Relatedly, I conclude that continued efforts to precisely pin down the natural rate of unemployment and to estimate its impact on current and near-term inflationary pressures are unlikely to be fruitful. In closing, I suggest we can more readily advance our ability to assess current and near-term inflationary pressures by developing better measures of expected inflation, and a deeper understanding of how expected inflation behaves and feeds into current inflationary pressures.

A DOWNWARD DRIFT IN THE TREND UNEMPLOYMENT RATE Movements in the U.S. unemployment rate are well approximated by the steady state relation, $u \approx u^{SS} = s/(s+f)$, where s is the monthly unemployment inflow rate and f is the monthly outflow rate. In light of this observation, Crump and colleagues' measure the trend unemployment rate, $\overline{u}$, by extracting trends in s and f. Although they use more disaggregated data in measuring the trend, their figure 1 tells the story: The inflow rate, s, drifts down from the early 1980s, falling by roughly half over nearly four decades. The outflow rate, f, is highly procyclical and shows some indication of a downward drift after 2000. On this basis, the authors conclude that the trend unemployment rate, $\overline{u} = \overline{s}/(\overline{s} + \overline{f})$, reflects a large downward drift in the unemployment inflow rate. The proportionally modest decline in $\overline{f}$ works in the opposite direction.

WHAT IS BEHIND THE DOWNWARD DRIFT IN $\overline{s}$? Crump and colleagues identify four factors behind the downward drift in $\overline{s}$ in recent decades: an increased labor force attachment of females, the aging of the U.S. population ("worker aging"), a rightward shift in the employment-weighted age distribution of firms ("firm aging"), and secular declines in job destruction and reallocation rates. As the authors rightly note, these factors overlap, and we cannot simply add them up to get their combined contribution.

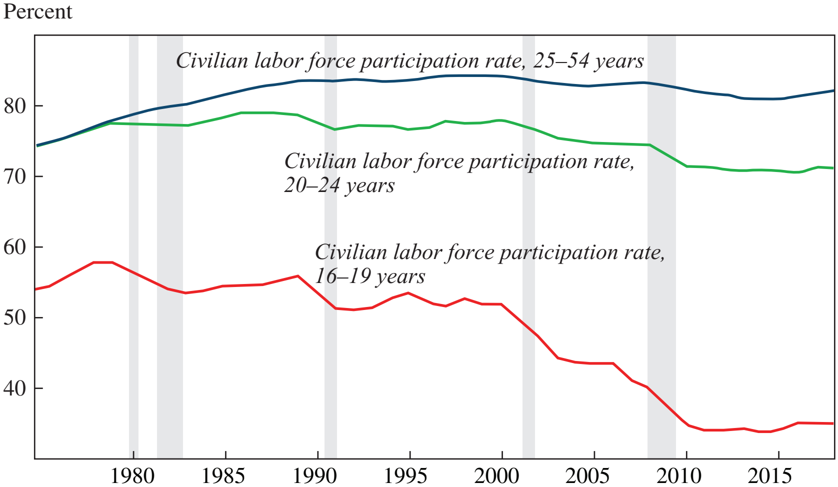
I agree that these four factors are important drivers of the downward drift in $\overline{s}$, but there is more to the story. Another important factor is the disappearance of short-duration employment relationships. Using data derived from administrative records, Henry Hyatt and James Spletzer (2017) show that more than half the drop in hiring and separation rates from 1996 to 2012 reflects a declining incidence of jobs that start and end in the same calendar quarter. They also find that the shifting composition of workers and employers accounts for only 22 percent of the declining incidence of short-duration employment relationships, mostly due to the aging of workers and firms.[1] This finding tells us that the disappearance of short-duration employment relationships is largely distinct from the first three factors that the authors stress. I suspect it is also largely distinct from the secular fall in job destruction and reallocation rates, given how job flows are measured.[2]

Another factor is falling labor participation rates among young adults. My figure 1 shows large declines since the late 1980s in the participation

1. Shifts by worker education and gender and by firm size and industry play a much smaller role or work in the opposite direction, according to Hyatt and Spletzer.

2. See Davis and Haltiwanger (1998) for an extended discussion of how worker flow measures (such as hiring and separation rates) relate to job flow measures.

Figure 1. U.S. Labor Force Participation Rates by Age Group, Annual, 1975–2018[a]



Sources: U.S. Bureau of Labor Statistics; Current Population Survey; FRED database of the Federal Reserve Bank of Saint Louis.

a. This figure shows data from the U.S. Bureau of Labor Statistics, specifically, the Civilian labor force participation rates from the Current Population Survey, retrieved from FRED on April 14, 2019: 16–19 years, LNU01300012; 20–24 years, LNU01300036; and 25 to 54 years, LNU01300060. Plotted data are annual averages of monthly values.

rates of persons who are 16–19 years old, more modest declines for those who are 20–24, and little change for those who are 25–54. My table 1 documents the well-known fact that young adults have relatively high unemployment rates. Unemployment inflow rates are also much higher among the young, as Crump and colleagues show in their figure 3. Taken together, these facts tell us that labor force participation rates by age shifted in a manner that contributes to the secular fall in $\overline{s}$ and $\overline{u}$. This factor is distinct from the role of population aging that the authors stress.

See my paper with John Haltiwanger (2015) for a discussion of developments that contributed to falling labor force participation rates among the young—and among the less-educated, who also have a relatively high propensity for unemployment. See Davis and others (2007), Decker and others (2014), and Davis and Haltiwanger (2015) on various factors behind the secular decline in job reallocation intensity.

THE SMALLER DOWNWARD DRIFT IN $\overline{f}$ The downward drift in the unemployment outflow rate is modest but has material implications for $\overline{u}$, as Crump and colleagues show. So, it is worth asking what deeper forces lie behind

Table 1. U.S. Civilian Unemployment Rates by Age Group and Time Period[a]

	Age group		
Time period	*25–54 years*	*20–24 years*	*16–19 years*
1975–2018	5.2	10.4	17.9
1975–99	5.2	10.5	17.8
2000–2018	5.1	10.2	18.1

Sources: U.S. Bureau of Labor Statistics; Current Population Survey; FRED database of the Federal Reserve Bank of Saint Louis.

a. This table shows data from the U.S. Bureau of Labor Statistics, specifically, the unemployment rates from the Current Population Survey, retrieved from FRED on April 14, 2019: 16–19 years, LNU04000012; 20–24 years, LNU04000036; and 25–54 years, LNS14000060. Table entries are averages of monthly values during the indicated period.

the downward drift in $\bar{f}$. In this respect, one noteworthy development is that geographic mobility has fallen in recent decades, even conditional on age (Molloy, Smith, and Wozniak 2014). As job losers and labor market entrants become less willing or able to migrate away from declining cities and regions, one likely effect is a fall in the unemployment outflow rate.

There are good reasons to think that falling geographic mobility is at least partly due to policy developments. As one example, the spread of occupational licensing (Kleiner and Krueger 2013) inhibits mobility across occupations and states. See the papers by Dick Carpenter and others (2012), the White House (2015), and Janna Johnson and Morris Kleiner (2017) for evidence. As a second example, Chang-Tai Hsieh and Enrico Moretti (2019) document a secular rise in the dispersion of nominal wages across U.S. cities from 1964 to 2009. They link this development to the adoption of land use restrictions in high-productivity coastal cities that reduced the elasticity of the housing supply and inhibited the in-migration of new workers from less productive cities and regions.

This raises interesting empirical questions that I have not seen addressed: To what extent is the downward drift in $\bar{f}$ concentrated in cities and regions with relatively low nominal wages? Is the downward drift present in cities with high nominal wages? Has the geographic dispersion of unemployment outflow rates risen in recent decades? If so, does the spatial pattern of rising dispersion in unemployment outflow rates align with the spatial pattern of rising nominal wage dispersion shown by Hsieh and Moretti (2019)?

FIVE TAKEAWAYS The authors make a compelling case that the trend component of the U.S. unemployment rate has fallen by roughly 3 percentage points since the early 1980s. They also identify several proximate drivers of this trend decline, and I have added to their list in the discussion above. These empirical results are quite helpful in assessing past U.S. economic performance and the future outlook for the U.S. economy, quite apart from any implications for inflation. In these respects, I see five important takeaways:

1. Much of the downward drift in U.S. unemployment rates over the past 35 years reflects good fortune rather than good macroeconomic policy. The "good fortune" includes the effects of population aging, increases in female labor force attachment, and declining business dynamism.

2. Some part of the downward drift in $\bar{u}$ probably reflects bad policy—that is, policies that drove younger, less educated, and other marginal workers out of the labor force.

3. Some past drivers of falling $\bar{s}$ have largely played out, and some may reverse. The increased labor force attachment of females, for example, seems to have largely played out. In all likelihood, we will be less fortunate with respect to the behavior of $\bar{s}$ and $\bar{u}$ in the coming decades.

4. The downward drift in $\bar{f}$, though modest and more recent, warrants concern. As the authors show (their figure 7, top panel), the fall in $\bar{f}$ raised $\bar{u}$ by 50 basis points. Lower values of $\bar{f}$ also slow recovery from the upward spikes in job destruction rates and s that typify the onset of recessions. That is, recessionary increases in unemployment take longer to unwind when $\bar{f}$ is lower, other things equal. Insofar as greater land use restrictions drove the downward drift in $\bar{f}$ by inhibiting migration to cities with better job opportunities, it will be politically challenging to reverse the decline in $\bar{f}$.

5. The trend component of the natural rate of unemployment has fallen substantially since the early 1980s, with 3 percentage points as a reasonable rough guess for the size of the fall.

THE CHALLENGE OF THE PHILLIPS CURVE Crump and colleagues integrate two very different approaches to estimating the natural rate of unemployment. This is a worthy ambition, and it is hard to take issue with the broad goal. Nevertheless, I see huge challenges in using the Phillips Curve to (1) sharpen our estimate of the natural rate of unemployment and (2) serve as a practical aid to monetary policymakers in assessing near-term inflationary pressures.

Justin Wolfers forcefully expresses one set of concerns about Phillips Curve modeling in his comments on a paper by Laurence Ball and Sandeep Mazumder (2011, 403–4) at a previous conference for the *Brookings Papers on Economic Activity*:

> That . . . the Phillips curve has not been proved false . . . may be because it is not falsifiable. . . . There are so many degrees of freedom to consider. . . . Inflation can be measured either as headline, core, or median, using either the PCE deflator,

> the CPI, or the GDP deflator. Inflation expectations can be modeled as rational, adaptive, or anchored. Data from different surveys can be utilized, such as the Livingston, the SPF, the Blue Chip, and the Michigan survey. Different measures of slack can be used, from the unemployment rate to the output gap to capacity utilization. The long-term unemployed can be included or not. Coefficients can be fixed or allowed to change over time. The lag structure can be adjusted, and nonlinearities can be assumed or ignored. Regime shifts can be invoked. Supply shocks can be included, including shocks to food, energy, and import prices, and price controls can be a factor in certain periods. Some economists in addition want to control for productivity or the labor share. In the end, there are more degrees of freedom than observations, which means that whatever path inflation might take, some researcher could plausibly claim to have found a Phillips curve that accounts for that path.

To this expansive list of Phillips curve variants, the paper at hand adds new degrees of freedom in the form of a richer, more flexible characterization of the natural rate of unemployment. Moreover, Crump and colleagues' account of puzzling U.S. inflation behavior during and after the global financial crisis leans very heavily on the paths of expected future inflation and expected future slack. We at least have multiple sources of data on expected future inflation, but expected future slack is essentially a free path variable, constrained only by the model and its functional form. Economists are very good at devising models with free path variables to fit nettlesome time series. From my vantage point, the authors' Phillips curve looks like the latest iteration in the long line of iterations that Wolfers summarizes. I do not think it will be the last iteration. More important, I do not see any reason to think it will prove a more useful practical guide to near-term inflationary pressures than many of its predecessors.

There is another view. At the same Brookings conference, James Stock responded to Wolfers as follows (Ball and Mazumder 2011, 404):

> The basic fact remain[s] that inflation in the United States and in other developed economies falls during periods of slack. This happened during the 1960s recession and again during the 1969 recession. The 1973 recession was different because of the oil price shock, but the pattern reappeared in the early-1980s and 1990 recessions, and again in 2000 for a while, except for a very interesting episode in 2004 and 2005. And much the same thing happened in 2007 and after, although the scale of it was in question. The issue then is not whether the pattern exists, but how to model it.

Stock's point about the "basic fact" is important and hard to deny. But it does not necessarily follow that the relationship between inflation and the unemployment gap (or other measures of slack) is sufficiently simple, stable, and predictable in its response to policy itself as to admit an empirical model that delivers confident predictions about near-term inflationary pressures.

In this respect, I am reminded of the view expressed by Olivier Blanchard (2016, 31): "Macroeconomists have learned, often painfully, that, while low unemployment creates inflation pressure, the form of the relation can change and has changed over time." Blanchard reviews some of these changes in U.S. unemployment-inflation dynamics since the 1960s.

In using the Phillips curve as a practical tool of monetary policy, perhaps the best we can do is to keep in mind Stock's "basic fact" and combine loose theorizing, simple statistical models, and informed judgment to obtain a very rough barometer of near-term inflationary pressures.

UNCERTAINTY ABOUT THE NATURAL RATE OF UNEMPLOYMENT The 68 percent confidence interval for u^* in the authors' preferred empirical model is about 1 percentage point. The 95 percent confidence interval is about 2 percentage points. Their figure 8 shows two estimated time series for u^*, one based on a price inflation Phillips curve, and one that relies on a Phillips curve specification that uses price and wage inflation data. As the authors discuss, their (median) estimated u^* differs a good deal between these two specifications during much of the 1970s and in the 2009–10 period. Eyeballing figure 8, it appears the peak difference is roughly 80 basis points. In addition to these sources of uncertainty about u^*, the estimated natural rate presumably depends on the choice of inflation expectations data to feed into the model, functional form choices, and more. Pulling these points together, I conclude that the authors' empirical undertaking does not yield much confidence about the value of the natural rate of unemployment at any point in time. Accounting for estimation uncertainty, specification uncertainty, and uncertainty about the appropriate data inputs, the range of reasonable values for u^* seems to be at least 250 basis points.

Here, I do not mean to suggest that the authors have done a poor job. Rather, I conclude that bringing the Phillips curve to the table helps little in sharpening our estimates for the natural rate of unemployment.

TAKING CRUMP AND COLLEAGUES' MODEL AT FACE VALUE: DOES SLACK MATTER? The authors estimate a flat Phillips Curve: the 90 percent confidence interval for the coefficient on the unemployment gap, $u_t - u_t^*$, is 0.011 to 0.031 in the model that uses price inflation data only and 0.018 to 0.041 in the model that uses both price and wage inflation data. Suppose these estimates are in the right ballpark. It follows that high uncertainty about the unemployment gap matters little for assessing current inflation pressures, so long as we have good data on expected inflation rates. To be concrete, suppose we misjudge the current value of u^* (and the unemployment gap) by 2 percentage points. Multiplying this misjudgment by a slope coefficient of 0.03 means that we misjudge current inflation pressures by only 6 basis points (annualized), conditional on expected future inflation. This is a tiny error. Indeed, it is probably smaller than the uncertainty about the current inflation rate.

The obvious corollary is that getting a sharp estimate for u^* matters very little for assessing current and near-term inflation pressures, provided that we have timely, high-quality measures of inflation expectations. In light of this corollary and 50 years of frustration in macroeconomists' efforts to develop a stable, reliable Phillips curve model, perhaps we should shift our focus to better measures of expected inflation, a deeper understanding of what causes expected inflation to move, and a better grasp on how expected future inflation feeds into current inflationary pressures.

REFERENCES FOR THE DAVIS COMMENT

Ball, Laurence, and Sandeep Mazumder. 2011. "Inflation Dynamics and the Great Recession." *Brookings Papers on Economic Activity*, Spring, 337–81; and comments and discussion, 382–405.

Blanchard, Olivier. 2016. "The Phillips Curve: Back to the 60s?" *American Economic Review* 106, no. 5: 31–34.

Carpenter, Dick M., Lisa Knepper, Angela C. Erickson, and John K. Ross. 2012. *License to Work: A National Study of Burdens from Occupational Licensing,* Arlington, Va.: Institute for Social Justice. https://ij.org/wp-content/uploads/2015/04/licensetowork1.pdf.

Davis, Steven J., and John C. Haltiwanger, 1998. "Measuring Gross Worker and Job Flows." In *Labor Statistics Measurement Issues*, edited by John Haltiwanger, Marilyn Manser, and Robert Topel. Chicago: University of Chicago Press.

———. 2015. "Labor Market Fluidity and Economic Performance." In *Economic Policy Symposium Proceedings: Re-Evaluating Labor Market Dynamics.* Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Davis, Steven J., John Haltiwanger, Ron Jarmin, and Javier Miranda. 2007. "Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms." In *NBER Macroeconomics Annual 2006,* edited by Daron Acemoglu, Kenneth Rogoff, and Michael Woodford. Cambridge, Mass.: MIT Press.

Decker, Ryan, John Haltiwanger, Ron S. Jarmin and Javier Miranda, 2014. "The Secular Decline in Business Dynamism in the U.S." Working paper. http://econweb.umd.edu/~haltiwan/DHJM_6_2_2014.pdf.

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation." *American Economic Journal: Macroeconomics* 11, no. 2: 1–39.

Hyatt, Henry R., and James R. Spletzer. 2017. "The Recent Decline of Single Quarter Jobs." *Labour Economics* 46: 166–76.

Johnson, Janna E., and Morris M. Kleiner. 2017. "Is Occupational Licensing a Barrier to Interstate Migration?" NBER Working Paper 24107. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24107.pdf.

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.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak, 2014. "Declining Migration within the U.S.: The Role of the Labor Market." NBER Working Paper 02138. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w20065.pdf.

White House. 2015. "Occupational Licensing: A Framework for Policy Makers." Report prepared by Council of Economic Advisers, Department of the Treasury, and Department of Labor. https://obamawhitehouse.archives.gov/sites/default/files/docs/licensing_report_final_nonembargo.pdf.

COMMENT BY

GIORGIO E. PRIMICERI The goal of this paper by Richard Crump, Stefano Eusepi, Marc Giannoni, and Ayşegül Şahin is to estimate the natural rate of unemployment, or u_t^*, in the postwar period. The authors combine two measurement approaches, one based on detailed data on flows into and out of unemployment for many demographic groups, and the other on the traditional Phillips curve relationship and data on aggregate unemployment, inflation, and inflation expectations. The paper provides three main takeaways. First, u_t^* is estimated to be about 4 percent in 2018. Second, u_t^* appears to have been trending down since the late 1980s. Third, this downward trend is due to the secular decline in the separation rate, which in turn was caused by the increased labor force attachment of females, the decline in job destruction and reallocation intensity, and the dual aging in the labor market of workers and firms. Overall, this paper is an impressive piece of work, with crucial policy implications. The most obvious of them is that the current low level of unemployment is roughly sustainable in terms of inflation, given that it is similar to the estimated natural rate, and the unemployment gap is thus close to zero.

My comments are organized around two main points. I first try to unpack the estimates of u_t^* presented by Crump and colleagues, to shed light on their essential drivers. I conclude that data on inflation expectations seem crucial for the measurement of u_t^* in the paper. In contrast, the detailed labor market flow data play a less central role for the measurement of u_t^*, although they are of course crucial for the interpretation of its secular trend. I then analyze the implications of this unpacking exercise for the New Keynesian Phillips curve, which seems in better shape than what some recent critics have suggested.

WHAT DRIVES CRUMP AND COLLEAGUES' ESTIMATE OF u_t^*? The authors' baseline estimate of u_t^* has three key features: (1) u_t^* has been trending down since the 1980s; (2) it is roughly equal to 4 percent in 2018; and (3) the uncertainty around its path is sizable but not overwhelming. My objective here is to understand what ingredients of the authors' complex empirical model are essential to these findings. To this end, I present a battery of u_t^* estimates, obtained using a sequence of progressively more complex models, the last of which corresponds to the authors' baseline model.

Model 1. The starting point of my unpacking exercise is the traditional backward-looking Phillips curve,

$$\pi_t - \pi^* = \gamma(\pi_{t-1} - \pi^*) - \kappa(u_t - u_t^*) + s_t. \tag{1}$$

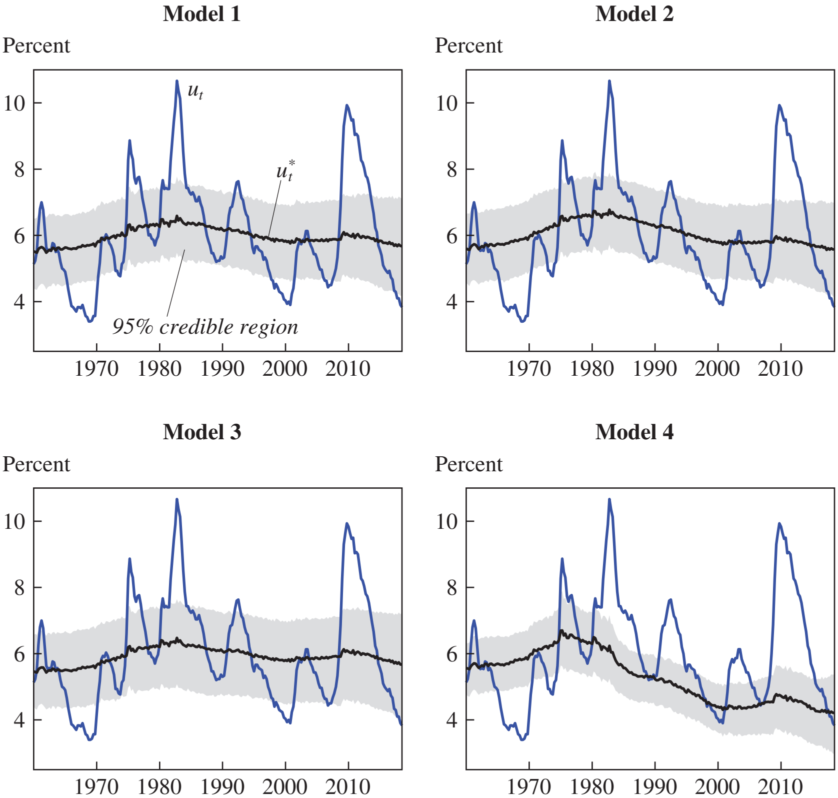
According to this "Triangle model" (Gordon 1977, 2013), deviations of inflation, π_t, from its target value, π^*, depend on an inertial term, a "demand factor" represented by the gap between unemployment, u_t, and its natural rate, u_t^*, and a supply shock, s_t. In the estimation, u_t^* is treated as an unobservable variable and modeled as a random walk process, to capture the idea that its movements are very persistent. Like Crump and colleagues, I assume that the unemployment gap and the supply shock follow a first- and second-order autoregressive (AR(2), AR(1)) process respectively, although these two assumptions are not crucial for the results. The top left panel of my figure 1 presents the implied estimate of u_t^*, which is quite different from the authors' baseline estimate: u_t^* is relatively stable over time, it is roughly equal to 6 percent in 2018, and the uncertainty around its path is large.

Model 2. I augment model 1 with all the ingredients of Crump and colleagues' baseline setup, adding these components one at a time to understand their specific roles. The first step in this direction consists of turning the backward-looking Phillips curve of equation 1 into a more modern, New Keynesian, forward-looking Phillips curve, based on sticky wages and indexation to past and steady-state inflation:

$$\pi_t - \pi^* = \gamma(\pi_{t-1} - \pi^*) - \kappa E_t \sum\nolimits_{j=0}^{\infty} \beta^j \left(u_{t+j} - u_{t+j}^*\right) + s_t. \tag{2}$$

The main new feature of equation 2 is that inflation depends on the expected present discounted value of the future unemployment gaps, and

Figure 1. Estimates of u_t^* Based on Models 1 , 2, 3, and 4, 1960–2019



Sources: U.S. Bureau of Labor Statistics; author's calculations.

not just on the current level of this gap, as in equation 1. However, this change has a relatively minor impact on the measurement of u_t^*, as evident from comparing the top right and top left panels of my figure 1.

Model 3. The next step in the direction of Crump and colleagues' model is to introduce time variation in the inflation target. More precisely, π^* in equation 2 is replaced by π_t^*, which is modeled as a random walk to capture the low-frequency, hump-shaped behavior of inflation during the 1970s and 1980s. The role of this modification, however, is also relatively marginal, resulting again in nearly unchanged estimates of u_t^* relative to models 1 and 2 (the bottom left panel of my figure 1).

Model 4. Moving on, I now estimate the model by also using data on short- and long-term inflation expectations. As with Crump and colleagues' estimate, these data are equated to the short- and long-term forecasts of inflation implied by the model, up to a measurement error. These survey data help pin down the level of the time-varying inflation target, π_t^*, among other things. Notice that the implied estimate of u_t^*, presented in the bottom right panel of my figure 1, is now quite different from the previous ones: u_t^* clearly trends down starting in the 1980s, it is close to 4 percent in 2018, and its uncertainty is lower. Overall, the path of u_t^* is quite similar to the authors' baseline estimate, suggesting that the use of data on inflation expectations is a crucial component of their empirical model.

Model 5. The final ingredient of Crump and colleagues' model is the assumption that u_t^* cannot permanently deviate from the secular trend of the unemployment rate, $\bar{u}_t$. In particular, the authors rewrite u_t^* as

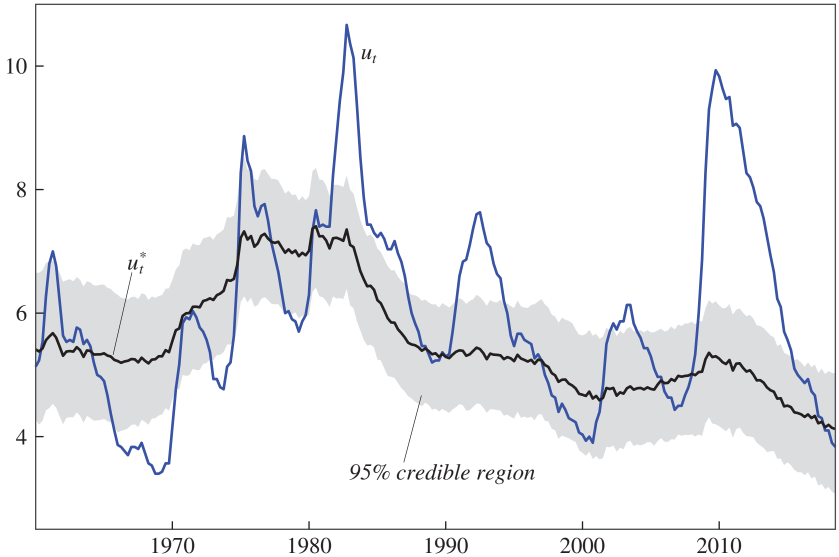
$$u_t^* = (u_t^* - \bar{u}_t) + \bar{u}_t,$$

where $\bar{u}_t$ is measured independently using disaggregated labor market flow data, and the term in parentheses—the distance between natural and secular unemployment—is assumed to follow an AR(1) process. I estimate this model using the authors' exact measure of $\bar{u}_t$, and I present the implied estimate of u_t^* in my figure 2. By construction, this estimate is identical to the authors' "inflation-only" estimate of u_t^*.

It is important to notice that the evolution of u_t^* shown in my figure 2 is similar overall to the path of u_t^* displayed in the bottom right panel of my figure 1. A possible interpretation of this finding is that disaggregated labor market flow data are not terribly useful to estimate u_t^*, because they do not drastically change our view about the time-series behavior of this variable. This interpretation, however, would probably be too literal and a bit naive. A more compelling view is that this consistency result—the fact that aggregate data on unemployment, inflation, and inflation expectations deliver estimates of natural unemployment in line with its secular trend—is remarkable, and provides an important external validation of the traditional Phillips curve framework.

IMPLICATIONS FOR THE NEW KEYNESIAN PHILLIPS CURVE The New Keynesian Phillips curve has recently been criticized because inflation fell little relative to the increase in unemployment during the Great Recession, the so-called missing disinflation phenomenon (Hall 2011). However—as stressed by Olivier Coibion and Yuriy Gorodnichenko (2015); Marco Del Negro, Marc Giannoni, and Frank Schorfheide (2016); and Carvalho and others (2017)—this somewhat puzzling behavior of inflation can be explained by the fact that inflation expectations were well anchored during the same

Figure 2. Estimate of u_t^* Based on Model 5, 1960–2019



Sources: U.S. Bureau of Labor Statistics; author's calculations.

period. By this logic, the explicit use of inflation survey data should robustify inference in the context of the New Keynesian Phillips curve, and improve the estimation of u_t^*, which is exactly what the previous empirical results show.

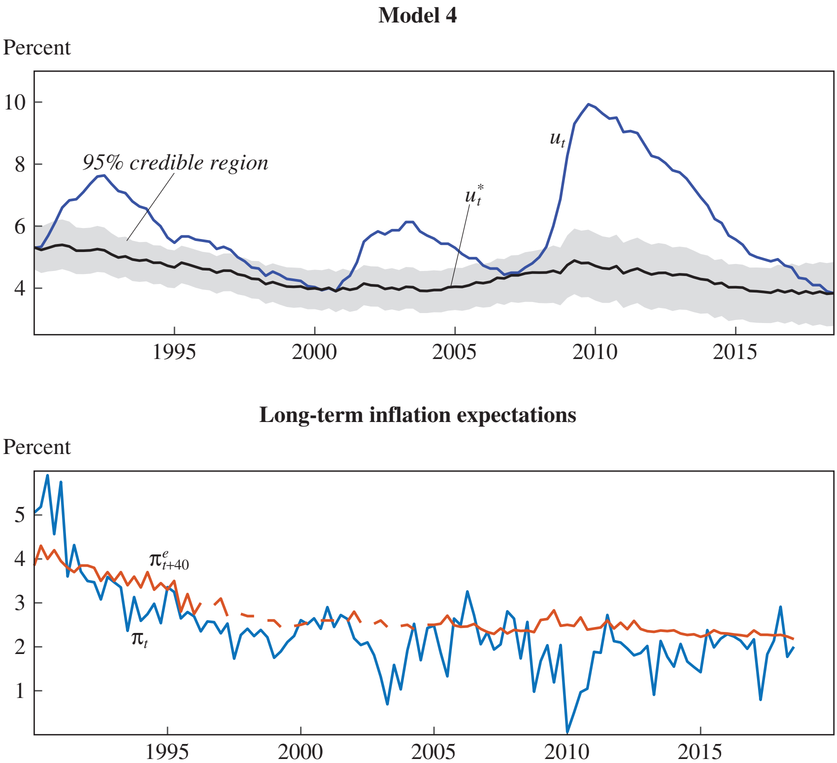
To understand why these survey data play such a crucial role, let π_t^* replace π^* in the New Keynesian Phillips curve given in equation 2, and rewrite this equation as

$$(3) \qquad \pi_t - \gamma\pi_{t-1} - (1-\gamma)\pi_t^* - \beta E_t[\pi_{t+1} - \gamma\pi_t - (1-\gamma)\pi_{t+1}^*] \\ = -\kappa(u_t - u_t^*) + \tilde{s}_t,$$

to make explicit the dependence of inflation on its expected future value.[1] The use of data on short- and long-term inflation expectations makes the variables $E_t\pi_{t+1}$ and π_t^* observable, up to some measurement error. As a consequence, it becomes easier for the econometrician to isolate the relationship between the left-hand side and the right-hand side of

1. The term $\tilde{s}_t$ is equal to $(1 - \beta\rho_s)s_t$, where ρ_s is the autocorrelation of the supply shock.

Figure 3. Post-1990 Estimate of u_t^* Based on Model 4 and Data on Inflation and Long-Term Inflation Expectations, 1990–2019

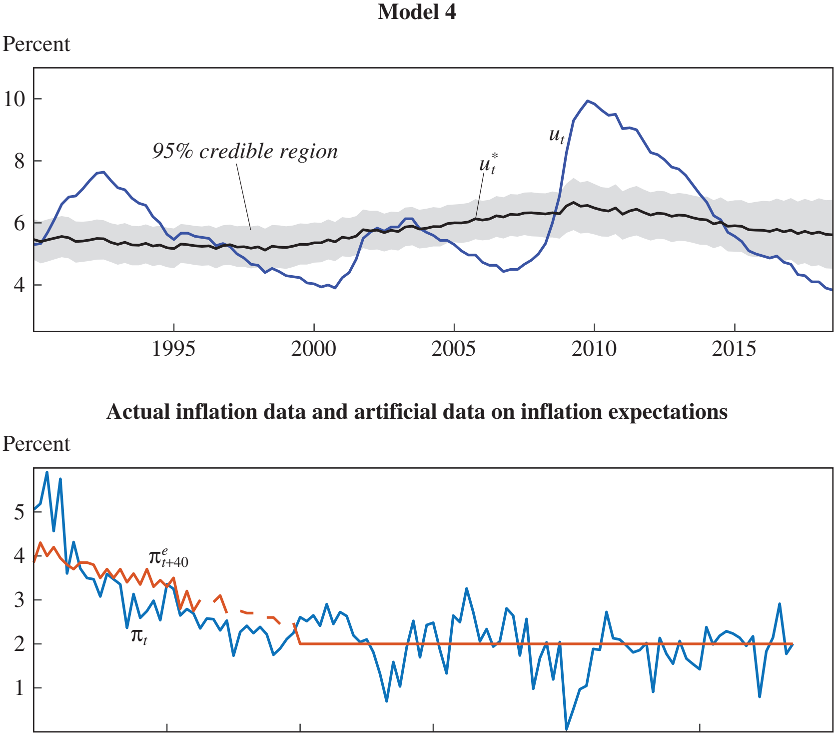


Sources: U.S. Bureau of Labor Statistics; Survey of Professional Forecasters; author's calculations.

equation 3, and to more precisely estimate the slope coefficients κ and the natural rate u_t^*. Furthermore, when using data on inflation expectations, inference about κ becomes also surprisingly stable over time. To illustrate this point, the top panel of my figure 3 presents the evolution of u_t^* according to model 4, when this model is estimated using only post-1990 data. Notice that the implied path of u_t^* is remarkably similar to the one based on the full-sample estimation of the same model, plotted in the lower right panel of my figure 1.

It is important to realize, however, that this powerful role of inflation expectation data in the estimation of the Phillips curve also comes with a disadvantage. The cost is the sensitivity of the estimate of u_t^* to the exact measurement of inflation expectations, which is notoriously difficult. For

Figure 4. Post-1990 Estimate of u_t^* Based on Model 4, Obtained Using Actual Inflation Data and Artificial Data on Inflation Expectations, 1990–2019



Sources: U.S. Bureau of Labor Statistics; Survey of Professional Forecasters; author's calculations.

example, the bottom panel of my figure 3 plots the time series of long-term inflation expectations used for the estimation of model 4 (and Crump and colleagues' baseline model). These expectations appear to be "upward biased," because agents systematically expect inflation in the long run to be higher than current inflation. Given that these survey data effectively pin down the level of π_t^*, this implies that actual inflation is almost always below target. In turn, this explains why the unemployment gap is almost always positive in the top panel of my figure 3.

To illustrate the sensitivity of u_t^* to the measurement of inflation expectations, the top panel of my figure 4 plots the implied evolution of u_t^* when I reestimate the model on post-1990 data, using a modified long-term inflation expectation series that is artificially set to 2 percent after 2000 (the bottom panel of my figure 4). The figure makes clear that the estimate of u_t^* would shift considerably in this counterfactual scenario, implying a substantially negative unemployment gap in 2018. This finding suggests that the Phillips curve is still relatively flat, despite being estimated quite robustly due to the explicit use of data on inflation expectations.

REFERENCES FOR THE PRIMICERI COMMENT

Carvalho, Carlos, Stefano Eusepi, Emmanuel Moench, and Bruce Preston. 2017. "Anchored Inflation Expectations." Working paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3018198.

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.

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.

Gordon, Robert J. 1977. "Can the Inflation of the 1970s Be Explained?" *Brookings Papers on Economic Activity*, no. 1: 253–77.

———. 2013. "The Phillips Curve Is Alive and Well: Inflation and the NAIRU during the Slow Recovery." NBER Working Paper 19390. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w19390.pdf.

Hall, Robert E. 2011. "The Long Slump." *American Economic Review* 101, no. 2: 431–69.

GENERAL DISCUSSION David Romer started the discussion by complimenting the paper and the commenters. He expressed the view, however, that the paper might have overemphasized u^* (the natural rate of unemployment) relative to $\bar{u}$ (the secular trend of unemployment).

One reason to use either of these two concepts is to forecast inflation, for which he argued u^* would clearly be the preferred measure. However, it may not matter much which variable is used because inflation does not seem to be reacting to labor market tightness very strongly.

Another reason to use either variable would be as a baseline for where the economy should move over the medium term. In this context, because u^* eventually reverts to $\bar{u}$, then $\bar{u}$ may be the better measure to focus on. For example, during the Great Recession, policies such as unemployment insurance raised u^* somewhat, while $\bar{u}$, a secular construct, remained mostly unchanged. Because monetary policymakers, for example, target economic outcomes over a longer horizon, $\bar{u}$ would have been a better metric to use. Similarly, based on the paper's estimates, unemployment is currently at about its natural rate (u^*), but is about 0.5 percentage point below its secular trend ($\overline{u}$). If u^* is likely to rise to $\overline{u}$ over the next several years, then again $\overline{u}$ would be the more relevant variable for policymakers to consider.

In addition, Romer agreed with the presenters about the relevance of inflation expectations in the modern version of the Phillips curve, but noted that there are two components to inflation expectations in modern models. One piece is essentially the present value of future output gaps (or minus the departures of unemployment from the natural rate). This piece has not been a very good predictor of inflation and behaved similarly during the Great Recession and during the Volcker-era recession. The second piece, however, is the long-term expectations of inflation. Romer expressed his belief that this piece is crucial to "saving" the Phillips curve, but that it lacks any substantive microeconomic foundations. Macroeconomic models get around this problem by assuming that workers index their wages to some abstract concept of expected inflation, but this assumption has little basis in reality. As a result, he argued that economists are still a long way from explaining why and how inflation expectations actually matter for inflation, and therefore from understanding the behavior of inflation.

Robert Hall complimented the paper on its litany of data on entrance and exit rates from unemployment. He noted though that the paper's conceptual reliance on the Phillips curve goes against data showing it to be an unreliable predictor of inflation. He cited a paper by James Stock and Mark Watson, as well as Phillips curves used in work by Robert Gordon as better examples.[1] These formulations of the Phillips curve study changes in unemployment relative to changes in inflation rather than the relationship between the level of unemployment and inflation and do a better job of fitting the data. He also referenced a paper he wrote with Thomas Sargent on the failure of the modern Phillips curve due to its misunderstanding of Milton Friedman's original ideas about the construct.[2] He noted specifically that Sargent had been critical of it since the 1960s.

1. James H. Stock and Mark W. Watson, "Modeling Inflation after the Crisis," *Proceedings: Economic Policy Symposium—Jackson Hole, Federal Reserve Bank of Kansas City*, 173–220, https://www.kansascityfed.org/publicat/sympos/2010/Stock-Watson_final.pdf; Robert J. Gordon, "The Phillips Curve Is Alive and Well: Inflation and the NAIRU during the Slow Recovery," NBER Working Paper 19390 (Cambridge, Mass.: National Bureau of Economic Research, 2013), https://www.nber.org/papers/w19390.pdf.

2. Robert E. Hall and Thomas J. Sargent, "Short-Run and Long-Run Effects of Milton Friedman's Presidential Address," *Journal of Economic Perspectives* 32, no. 1 (2018): 121–34, https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.32.1.121.

Frederic Mishkin noted that good monetary policy that acts to stabilize inflation will induce a bias toward estimating a flatter slope of the Phillips curve than would otherwise be the case. When inflation goes up, a responsible central bank should raise interest rates, which will increase unemployment. This effect would partially offset the negative slope between inflation and unemployment contained in the structural construct of the Phillips curve. He argued that the flattening of the Phillips curve's slope since the 1980s can be explained by this phenomenon and may bias estimates of the curve. Although the structural Phillips curve may still be active, monetary policy could be masking its empirical identification. He argued this was an important issue to address, particularly because a policymaker such as National Economic Council director Larry Kudlow seemed to be implying that the Phillips curve was dead and therefore the economy could sustain low interest rates without suffering from inflation.[3] Ignoring these types of issues could result in bad policy results, he argued.

Robert Gordon complimented the paper and its commenters, and he stressed the importance for policy of the paper's estimate of the natural rate of unemployment. In particular, if the paper's estimates were correct, then the economy has been at full employment for the past year and has been fluctuating around it for even longer, which explains why there has been so little inflation. An alternative explanation for low inflation would be that the natural rate is higher, maybe 5 percent, but that the Phillips curve is very flat. The distinction between the two narratives is important for monetary policy, he argued, because pushing the unemployment rate down to 3 percent would lead to inflation in the first narrative but not the second. The paper comes out on the side of the first argument. To understand which argument is correct, Gordon referenced the period between 2009 and 2015, when unemployment increased but there was little change in inflation, a point that would argue in favor of the second narrative, that the Phillips curve is very flat and there is little actual evidence for what the natural rate of unemployment is.

Picking up on comments from Romer and from Giorgio Primiceri in his comment, Gordon commented on the increasing importance of inflation expectations in the modern Phillips curve. He noted that the relevance of expectations may have come about as a result of the Federal Reserve's explicit inflation target or as some other process that stabilized inflation.

3. Larry Kudlow and Stephen Moore, "Who's Afraid of Higher Wages?" *Wall Street Journal*, March 12, 2018, https://www.wsj.com/articles/whos-afraid-of-higher-wages-1520897733.

Including inflation expectations in the Phillips curve causes an increase in estimates of the natural rate of unemployment in the 1970s and 1980s, from about 6 percent to around 8 percent, something that did not occur in Gordon's own work, which did not model the Phillips curve using inflation expectations. He noted that this difference was due to the fact that his own models explicitly include shocks specific to the time, rather than inflation expectations more broadly.

Finally, Gordon commented on interesting trends shown in the paper's data on labor market entry and exit from unemployment. He noted the mounting pile of evidence on the declining rate of entry of workers into unemployment and the declining importance of layoffs. As evidence of this, he noted the similarity in the levels of new claims for unemployment insurance today compared with the 1960s. The levels are comparable today and are still quoted in thousands, even though the size of the labor force has doubled since the 1960s. This point further indicates how dramatically entry into unemployment has declined.

Wendy Edelberg compared estimates of the natural rate of unemployment in the paper with estimates from the Congressional Budget Office (CBO). She inquired as to how the authors thought about labor force participation when making their estimates, wondering if differences in the treatment of this variable could drive divergence in the paper's estimate of u^* and the CBO's estimates. Specifically, though the CBO considers the inflationary pressures arising from the Phillips curve when estimating the natural rate, they weight the construct referred to as $\bar{u}$ in the paper more heavily because it takes into account longer-term structural trends like participation. Instead, the CBO had started to focus on the concept of an employment gap rather than an unemployment gap. Estimates of $\bar{u}$ in the paper are indeed similar to CBO's estimate of the natural rate of unemployment, showing values of about 4.6 percent in recent data. She asked how the authors calibrated their measures of u^* to measures of slack in the labor market. She noted that based on rough calculations, the labor market slack implied by the paper shows a similar unemployment gap in 2016 to the CBO's estimate of labor market slack after incorporating weakness in labor force participation relative to its potential. She stressed the importance of thinking about participation when estimating u^*.

Justin Wolfers expressed his frustration in explaining the Phillips curve to undergraduate economics students. Despite its centrality to macroeconomics, there are scant empirical illustrations of its existence. He joked that Paul Krugman's economic textbook solves this problem by only showing data on the Phillips curve from 1955 to 1968; N. Gregory Mankiw's textbook only shows data from 1961 to 1968, and Ben Bernanke's textbook shows no data at all (Bernanke agreed and laughed in response).

Wolfers also inquired about the confidence intervals around u^* in the paper. He interpreted the paper to estimate a very small probability that u^* was below 5½ percent in 1973. At Brookings in 1973, Arthur Okun presented a paper estimating the natural rate of unemployment to be between 4 and 5 percent, which the current paper views as statistically unlikely, indicating the uncertainty about estimates of u^* through time.[4]

Gerald Cohen argued that despite the centrality of inflation expectations in many modern versions of the Phillips curve, most economists are often overly precise about estimates of inflation expectations. He referenced comments from Primiceri, for example, noting that many survey respondents expect inflation to be about 2.5 percent rather than the Federal Reserve's 2 percent target over time. Cohen referenced a 2015 *Brookings Paper* studying inflation expectations in New Zealand, which showed that they were well above the reality of realized inflation as well as of the central bank's inflation target.[5] Therefore, estimating the Phillips curve with some precise estimate of inflation expectations, possibly based on forecasts from the Blue-Chip Economic Indicators survey filled out by informed professionals, may be detached from inflation expectations in reality.

Laurence Meyer gave his interpretation of the paper, notably that the Phillips curve is operative but not particularly relevant, or, as he put it, "The Phillips curve is alive, but who the hell cares?" In particular, the Phillips curve may exist, but it takes massive declines in unemployment to get small increases in inflation because the curve is so flat, meaning that in practice there is very little trade-off. Meyer noted that the paper goes on to argue that inflation expectations are instead central to the inflation process, and to get inflation, central bankers have to change the public's expectations of future inflation. Meyer argued, however, that this is an incredibly difficult task and represents a major crisis in central banking; central banks cannot run tight labor markets or change expectations to get inflation up to target. Both Japan and Europe, for example, are still a long way from

4. Arthur M. Okun, "Upward Mobility in a High-Pressure Economy," *Brookings Papers on Economic Activity*, no. 1 (1973): 207–52, https://www.brookings.edu/wp-content/uploads/1973/01/1973a_bpea_okun_fellner_greenspan.pdf. In this paper, Okun actually says that "at the present time, the controversial range for the target unemployment rate extends from 4 to 5 percent."

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

achieving their inflation targets on a sustained basis, despite significant forward guidance designed to increase inflation expectations as well as accommodative monetary policy driving tight labor markets. Under this interpretation, estimates of the natural rate of unemployment essentially seem irrelevant, given the flatness of the Phillips curve and how immovable inflation expectations are. The United States, however, appears to be in an advantageous position in this regard, given that inflation is close to target and inflation expectations seem well anchored. As Meyer put it, "We are just beautiful. But everybody else is screwed."

John Haltiwanger commented on the paper's dual approach to measure u^* using labor market entry and exit from unemployment and estimates of the Phillips curve. The former in particular involves measures of labor market tightness, which is difficult to measure. Early versions of labor market tightness involved vacancies over unemployment, but more recent versions involve modeling a broader job matching process. New models might measure slack as effective vacancies over effective searches, where effective searches include people out of the labor force as well those in unemployment. He cited Robert Hall and Sam Schulhofer-Wohl's recent paper showing that the probability of getting a job for someone who wants one but has fallen out of the labor force is relatively high and is also highly cyclical.[6] Measures such as these, which better measure labor market slack, might be better than u^* or $\overline{u}$ and might be used to provide a better fit of the Phillips curve.

Haltiwanger also commented on the suggested relationship in the paper between declining inflows into unemployment and declining labor market dynamism and reallocation. Sympathetic with this idea, he remarked that there might be a decline in the volatility of idiosyncratic labor market shocks that would be consistent with this trend, perhaps driven by changes in the composition of firms in the economy. However, the evidence has been pushing against this idea. Notably, new labor productivity data at the firm level in the United States shows an increasing dispersion of productivity growth across firms, a trend consistent with *rising* labor market frictions or wedges—a worrying sign.

Richard Cooper, echoing similar comments from Justin Wolfers, remarked on how different the paper's estimates of u^* were from historical estimates. Specifically, the *1962 Economic Report of the President*, which was written

6. Robert E. Hall and Sam Schulhofer-Wohl, "Measuring Job-Finding Rates and Matching Efficiency with Heterogeneous Job-Seekers," *American Economic Journal: Macroeconomics* 10, no. 1 (2018): 1–32, https://web.stanford.edu/~rehall/JobFindingRates.

by economic luminaries such as James Tobin and Arthur Okun, targeted an interim unemployment rate of 4 percent, an estimate below the 95 percent confidence interval in the paper.[7] He asked where historical estimates went wrong, questioning the paper's estimates of u^* for the early 1960s. Robert Gordon countered that the evidence was in favor of the paper's estimates because inflation was high in the 1960s as a result of underestimating u^* at the time.

Martin Baily reviewed generally the concept of u^* as originally conceived by Milton Friedman, noting that the paper treated u^* in the same spirit.[8] Namely, under Friedman's original framework, the unemployment rate falls below its natural rate accidentally. This is due to the fact that monetary policy affects inflation, which in turn drives down real wages. Because of downward nominal wage rigidity, firms tend to lay off workers instead of cutting their wages, hence leading to temporary surges in unemployment where the unemployment rate deviates from its natural rate absent wage rigidity. However, this basic framework may have shifted, primarily due to structural changes in the labor market institutions governing employment specifically related to weaker worker bargaining power and the decline of unions; making the wage setting process different could change the degree to which nominal wage rigidity binds. In addition, Baily reinforced the point made by David Romer: that by saying the Phillips curve has been "rescued" by including inflation expectations without a real theory behind how inflation expectations form, the authors' argument in favor of the Phillips curve leaves something to be desired.

George Perry echoed the point made by Martin Baily about labor unions. In the first 30 years after World War II, strong labor unions that could negotiate wages and salaries made wages more reactive to labor market conditions. This could give way to wage price spirals and inflation, making inflation more sensitive to labor market conditions and economic shocks more generally. During this period, the Phillips curve could be estimated well. The decline of labor unions since then, however, has instead led to a structural change in this relationship, Perry argued, and trying to find alternative specifications for the Phillips curve ignores these structural changes. Namely, the change came relatively abruptly after the 1980s, driven

7. Council of Economic Advisers, *1962 Economic Report of the President* (Washington: U.S. Government Printing Office, 1962), 8, https://fraser.stlouisfed.org/title/45/item/8133.

8. Milton Friedman, "The Role of Monetary Policy," speech given at Eightieth Annual Meeting of the American Economic Association and then published in the *American Economic Review* 58, no. 1 (1968): 1–17.

by foreign competition that weakened the bargaining power of domestic labor unions. Foreign competition also put general downward pressure on wages and prices. Trying to fit the same model to these two distinct periods invites policy mistakes. For example, current monetary policy—namely, inflation targeting—assumes that some modest amount of inflation can maximize employment regardless of the institutional setup of wage and price setting.

Stefano Eusepi thanked the discussants for their observations, promising to incorporate them into the paper. First, he addressed critiques of the Phillips curve by making a distinction about how exactly it has flattened in recent years. Namely, in their model, with the same size of the output gap, they can explain both the low inflation since 2008 and the high inflation of the 1970s. The reason for this is that the Phillips curve has only flattened in one dimension: its relationship with the current unemployment gap. It has *not* flattened in relation to the present discounted value of future unemployment gaps. This creates a situation where even if the unemployment gap is large, but it is expected to revert back to zero, then the Phillips curve is quite flat and inflation does not react. However, if the gap is expected to be permanent, then the slope of the Phillips curve can be quite large and inflation does materialize. In this way, the Phillips curve is not actually flat.

Eusepi also commented on the formation of inflation expectations. Although their current paper does not model inflation expectations, he noted other papers the authors had written on the behavior of long-term inflation expectations that vary endogenously with monetary policy.[9] These models can explain inflation expectations based on survey data, both in the 1970s and today.

James Stock referenced comments from Frederic Mishkin and Laurence Meyer on the flat Phillips curve. He asked about how inflation expectations are developed and their relationship with monetary policy. In this context, if the unemployment rate were expected to increase by a large amount, then monetary policy should respond by a large amount. If this is the case, then expectations should not move as much.

Ayşegül Şahin discussed the difference between u^* and $\bar{u}$, noting the usefulness of both measures. She referred to $\bar{u}$ as an anchor of u^*, and she cited the Great Recession as a prominent example. During the Great

9. Carlos Carvalho, Stefano Eusepi, Emanuel Moench, and Bruce Preston, "Anchored Inflation Expectations," working paper, https://centridiricerca.unicatt.it/complexity-Eusepi_AnchoredDraft_Final_Final.pdf.

Recession, when unemployment was high, $\overline{u}$ was trending down, due to secular factors. Therefore, recognizing this downward trend in $\overline{u}$ would have increased policymakers' confidence that there was substantial slack in the labor market.

In terms of bargaining power, Şahin referenced work by Andrew Glover and Jacob Short showing that aging is affecting the labor share of income because older workers extract less of the profit generated by firms.[10] The aging of the population is consistent with the decline in bargaining power, which also explains the decline in the inflow rate to unemployment discussed in the paper.

10. Andrew Glover and Jacob Short, "Demographic Origins of the Decline in Labor's Share," working paper, 2018, https://www.cemfi.es/ftp/pdf/papers/Seminar/shortj-gs2018_lsage.pdf.

CHRISTINA D. ROMER
University of California, Berkeley

DAVID H. ROMER
University of California, Berkeley

Fiscal Space and the Aftermath of Financial Crises: How It Matters and Why

ABSTRACT In a sample of 30 countries during the period 1980–2017, those with lower debt-to-GDP ratios responded to financial distress with much more expansionary fiscal policy and suffered much less severe aftermaths. Two lines of evidence together suggest that the relationship between the debt ratio and the policy response is driven partly by problems with sovereign market access, but even more so by the choices made by domestic and international policymakers. First, although there is some relationship between more direct measures of market access and the fiscal response to distress, incorporating the direct measures attenuates only slightly the link between the debt ratio and the policy response. Second, contemporaneous accounts of the policymaking process in episodes of major financial distress show a number of cases where shifts to austerity were driven by problems with market access, but show at least as many where the shifts resulted from policymakers' choices despite an absence of difficulties with market access. These results point to a twofold message: conducting policy in normal times to maintain fiscal space provides valuable insurance in the event of a financial crisis, and domestic and international policymakers should not let debt ratios unnecessarily determine the response to a crisis.

Conflict of Interest Disclosure: The authors did not receive financial support from any firm or person for 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.

There is enormous variation in macroeconomic performance in the aftermath of financial crises. Recent research finds that the amount of fiscal space countries have before a crisis—that is, the room policymakers have to take action—appears to be an important source of this variation. Countries that have low debt-to-GDP ratios when a crisis strikes typically face only modest downturns, while countries that have high debt ratios generally suffer large and long-lasting output losses (Jordà, Schularick, and Taylor 2016; Romer and Romer 2018). The apparent mechanism behind this correlation is the obvious one: countries that begin a crisis with ample fiscal space take much more aggressive fiscal action. This includes both financial rescue—bank bailouts, loan and deposit guarantees, and recapitalization of financial institutions—and conventional fiscal stimulus—tax cuts and spending increases (Romer and Romer 2018).

Our primary goal in this paper is to understand *why* a country's fiscal response to a crisis depends on its prior debt-to-GDP ratio. One possibility is that it reflects constraints imposed by market access. Countries with a higher debt ratio may be less able to take aggressive fiscal action or must move more quickly to austerity than lower-debt countries because investors push sovereign yields to prohibitive levels or refuse to lend to them entirely. Alternatively, the link between the fiscal response to a crisis and a country's debt-to-GDP ratio may reflect choices made by the country or by international organizations. For example, policymakers' ideas may lead them to tighten fiscal policy after a crisis if the debt ratio is high, but not otherwise. Likewise, the views of international organizations, such as the European Union and the International Monetary Fund, may be tied to the debt ratio, and may drive fiscal policy after a crisis either indirectly (say, through standing EU rules) or directly (through bailout conditionality).

We investigate this issue using both statistical and narrative evidence for the period since 1980 for 30 countries that belong to the Organization for Economic Cooperation and Development (OECD). Our finding is that both market access and policymakers' choices have played important roles in the fiscal response to crises over the past 40 years, but choices have been somewhat more central.

A crucial input into our analysis is the indicator of financial distress derived from narrative documents for 24 OECD countries described in our 2017 paper. Here, we extend this indicator through 2017 and incorporate the 6 countries that joined the OECD between 1973 and 2000. We thereby increase the number of observations covered by our measure by more than 20 percent, and the number where our measure shows positive levels of distress by 50 percent. In addition, the inclusion of countries such as Mexico, South Korea, and Hungary allows us to see if less advanced economies fare differently after crises than more mature ones. Extending the series through 2017 allows us to do a much more complete analysis of the aftermath of the 2008 global financial crisis than was possible in our previous study, which ended in 2012. For the most part, we find that the extended series yields results similar to those in our previous paper. The average aftermath of a crisis remains negative, highly persistent, and of moderate severity. Contrary to what one might expect, the aftermath of a crisis is somewhat less severe on average in less advanced economies. Consistent with our previous study, we also find that there is tremendous variation in the aftermaths of crises. Indeed, if anything, including a wider range of countries and more years after the global financial crisis makes the variation even starker.

To document the importance of fiscal space for the aftermath of crises and the fiscal response, we run panel regressions of output and the high-employment surplus at various horizons after time *t* on financial distress at *t*, including an interaction between distress and the prior debt-to-GDP ratio. The coefficient on the interaction term is consistently highly significant and of the expected sign: high-debt countries have larger output losses after a crisis and undertake fiscal contraction rather than expansion. The extensive literature on the impact of tax changes and government spending on output suggests that there is likely a causal relationship between these two developments. Likewise, focusing on the 22 episodes of high financial distress in our sample confirms a strong correlation between the size of the fiscal expansion after a crisis and the prior debt-to-GDP ratio.

The possibility that the debt ratio matters for the fiscal response to crises because it affects sovereign market access (or because it proxies for market access) can be investigated empirically. Interest rates on government debt, sovereign credit default swap (CDS) spreads, and credit-agency ratings are all direct indicators of market access. Likewise, being subject to a bailout program from the IMF or another international institution likely reflects severe problems with obtaining sovereign funding in private markets. If a country's debt-to-GDP ratio affects its fiscal response to a crisis through market constraints, including such direct measures of market access (interacted with financial distress) in the panel regressions should greatly weaken or eliminate the predictive power of debt for the fiscal response. It does not. Although some of the direct measures of market access do seem to affect the fiscal response to a crisis, the effects are generally moderate and are only marginally significant. At the same time, the interaction effect with the debt ratio remains significant and quantitatively important when the direct measures of market access are included. That is, countries with little fiscal space as measured by their debt-to-GDP ratio undertake less fiscal expansion after a crisis than their lower-debt counterparts, even controlling for the interest rates on their debt and other obvious indicators of market access. This supports the view that choices play an important role in countries' fiscal decisions during and after crises.

More evidence on the nature and determinants of the fiscal response to crises can be obtained from narrative sources. In particular, we read the *Country Reports* from the Economist Intelligence Unit (EIU) for the four years after the start of high financial distress in the 22 crisis episodes in our sample. The EIU's reports provide a blend of political and policy information that is particularly useful for deducing the motivation for fiscal actions during and after financial crises. A systematic reading of the reports shows that in some cases, problems with market access unquestionably led to fiscal contraction despite severe postcrisis recessions. This was the case, for example, in Spain and Italy after the 2008 global financial crisis. Sometimes severe market access problems led to an international bailout, with the result that fiscal policy in the affected countries was then driven partly by the views of the rescuing organizations; this was the case, for example, with Mexico after its crisis in the mid-1990s and with Portugal and Greece after the global financial crisis. In many other cases, however, the EIU suggests that the fiscal response to a crisis was driven by the choices of domestic policymakers, and, in the case of some EU countries, by EU rules and ideas. This is always true of postcrisis fiscal expansions, which are inherently discretionary. But choices were also often central to postcrisis austerity, such as that in the United Kingdom and Austria after the global financial crisis. Indeed, in roughly half the cases of postcrisis fiscal austerity, the EIU indicates that policymakers' ideas were more important than market access. The EIU's *Country Reports* also provide substantial narrative evidence that both market access and policymakers' choices were related to the debt-to-GDP ratio.

Our analysis of the role of fiscal space in the aftermath of financial crises is organized as follows. Section I discusses the extension of our narrative measure of financial distress, and revisits our basic findings about the average aftermath of a financial crisis and the variation in outcomes. Section II presents statistical results on the role of the debt-to-GDP ratio in explaining the variation in the aftermaths of crises. Section III discusses quantitative evidence on whether the debt-to-GDP ratio matters for the fiscal response to crises because it works through or proxies for market access. Section IV provides narrative evidence on the determinants of the fiscal response after a financial crisis. Finally, section V presents our conclusions and discusses the implications of our findings for economic policy.

Our study builds on several lines of research. First, it is obviously related to the large, but differently focused, literature on the aftermath of financial crises (for example, Bordo and others 2001; Reinhart and Rogoff 2009; Romer and Romer 2017; Baron, Verner, and Xiong 2019). Second, such authors as Henning Bohn (1998), Enrique Mendoza and Jonathan Ostry (2008), and Atish Ghosh and others (2013) investigate how the conduct of fiscal policy varies with the debt-to-GDP ratio. These papers, however, do not address either how the debt ratio affects the fiscal policy response to financial crises or the mechanisms through which the debt ratio affects the conduct of policy. Third, work defining and measuring fiscal space (for example, Ghosh and others 2013; Kose and others 2017) is also somewhat relevant to the issues we study. Relatedly, Maurice Obstfeld (2013), Douglas Elmendorf (2016), and other observers argue that having greater fiscal space can be very valuable in the event of a financial crisis. Our analysis lends strong support to this view.

Our research is clearly also related to the voluminous literature on the output effects of fiscal policy (for a recent survey, see Ramey 2016). The subset of this literature that examines whether fiscal multipliers are larger when the debt-to-GDP ratio is lower (for example, Perotti 1999; Ilzetzki, Mendoza, and Végh 2013) is closer to the issues we address. However, our finding that the fiscal policy response to financial crises is expansionary at low debt ratios and contractionary at high debt ratios means that the mechanism through which debt affects outcomes in our analysis is different than in those papers.

Finally, the two papers most closely related to our contribution here are the one by Òscar Jordà, Moritz Schularick, and Alan Taylor (2016) and our 2018 paper. Both find that the aftermath of a financial crisis is far worse in countries with high levels of government debt, and our 2018 paper finds that a likely mechanism behind this link is that the policy response is far more contractionary in high-debt countries.[1] One contribution of this paper is to extend and amplify these findings. But our main focus, which

1. Bernardini and Forni (2018) extend this analysis to consider both the level of government debt and its rate of change. They find that when both variables are unusually high before a recession that is associated with a financial crisis, the recession is unusually severe, and real per capita government spending falls rather than rises. They also show that reliance on IMF credit rises more than usual in such cases, which is suggestive of problems with market access.

these papers do not address, is on the reasons for the dependence of the policy response on the level of debt.

I. Preliminaries

In order to analyze the aftermath of financial crises, one needs a reliable indicator of when crises have occurred in various countries. We begin with the scaled index of financial distress in 24 OECD member countries derived from narrative records described in our 2017 paper. For this paper, we extend the index through 2017 and incorporate 6 additional countries. This section describes this extension and briefly discusses its impact on some of our previous results.

I.A. Extending the Measure of Financial Distress

Our measure of financial distress has three defining characteristics. One is that it is derived from contemporaneous narrative sources. In particular, it is based on the *OECD Economic Outlook*, a semiannual review of economic and financial conditions in each OECD country. Because the *Economic Outlook* is available beginning in 1967, our series on financial distress also begins then. There are two observations per year (corresponding to the two issues of the *Economic Outlook*), dated approximately June and December. Throughout the paper, we use the notations "H1" and "H2" to denote the two halves of the year.

Second, we take as our definition of financial distress Ben Bernanke's (1983) concept of a rise in the cost of credit intermediation—that is, something causes it to be more costly for financial institutions to supply credit at a given level of the safe interest rate. This could be an increased external cost of funds due to a widespread loss of confidence; increased costs of monitoring borrowers; or an increased internal cost of funds because of rising loan defaults.

Third, we scale financial distress along a continuum. This reflects the reality that, like most things, financial distress is not a 0/1 variable. To do this, we define our measure from 0 (no distress) to 15 (extreme crisis; widespread chaos and paralysis in the financial system). Values of 7 and above roughly correspond to what the IMF and the creators of other chronologies would identify as a systemic financial crisis (Laeven and Valencia 2014). In our analysis, we therefore often pay particular attention to episodes where distress reached 7 or more.

To construct our measure, we specify detailed criteria for translating OECD analysts' words into our numerical scale. Because the OECD does not typically talk in terms of the cost of credit intermediation, this involves looking for sensible proxies in the narrative accounts. Does the *Economic Outlook* discuss funding difficulties for banks, a breakdown in intermediation, or creditworthy borrowers having difficulty getting loans? Does it describe the problems as relatively minor (or perhaps as affecting just a small sector of the economy), or as severe and widespread? Does it believe that troubles in the banking system are just a risk to the forecast, or central to the outlook? In online appendix A of our 2017 paper, we describe the criteria for different levels of distress in detail, and we provide a summary of the reasoning (and the related quotations from the *OECD Economic Outlook*) for the observations we scale greater than zero.

Our original index covered the period 1967–2012. We also limited our analysis to the 24 countries in the OECD as of 1973. For this paper, we continue the narrative analysis through 2017. We also add the 6 countries that joined the OECD between 1973 and 2000: the Czech Republic, Hungary, Mexico, Poland, the Slovak Republic, and South Korea.

We use the same criteria and approach as we did for our original study. The one difference is that previously, we used key word searches (for terms such as "crisis" and "bank") to narrow down the number of country entries we needed to read word for word. Because most countries were still recovering from the 2008 global financial crisis between 2012 and 2017, we found it simpler to just read every entry for this period. Likewise, for the added countries, we felt it prudent to read all their entries in the *Economic Outlook* because some (particularly the former communist countries) only gradually developed the market-based financial systems that fit into our classification system. For these countries, we do not define our measure of financial distress until the descriptions in the *Economic Outlook* make clear that the country's financial system was largely privatized, and that its credit availability was therefore mainly determined by market forces.[2]

Table 1 shows the nonzero values of our measure of financial distress for all 30 countries for the period 2013:H1–2017:H2. It also shows all the

2. The starting dates for our measure for these countries are 2003:H2 for the Czech Republic (which joined the OECD in December 1995 and first appeared in the 1996:H1 issue of the *Economic Outlook*); 1998:H1 for Hungary (May 1996 and 1996:H1); 1998:H1 for Poland (November 1996 and 1996:H2); and 2003:H2 for the Slovak Republic (December 2000 and 2000:H2). For Mexico and South Korea, we define our measure starting when they first appeared in the *Economic Outlook* (1994:H1 for Mexico and 1996:H2 for South Korea). Online appendix A discusses the narrative evidence for the appropriate starting date for the added countries.

Table 1. Financial Distress in Countries Belonging to the Organization for Economic Cooperation and Development, 2013:H1–2017:H2 (and in the Added Countries Starting When Information Is Available)[a]

Austria
2013:H1 Credit disrupt.–reg.
2013:H2 Credit disrupt.–reg.
2014:H1 Credit disrupt.–reg.
2014:H2 Credit disrupt.–minus
2015:H1 Credit disrupt.–minus
2015:H2 Credit disrupt.–minus
2016:H1 Credit disrupt.–minus

Czech Republic
2008:H1 Credit disrupt.–minus
2010:H1 Credit disrupt.–minus

Denmark
2013:H1 Credit disrupt.–minus
2013:H2 Credit disrupt.–minus

France
2013:H2 Credit disrupt.–minus

Germany
2013:H1 Credit disrupt.–minus
2014:H1 Credit disrupt.–minus

Greece
2013:H1 Moderate crisis–minus
2013:H2 Minor crisis–plus
2014:H1 Minor crisis–reg.
2014:H2 Minor crisis–reg.
2015:H1 Moderate crisis–reg.
2015:H2 Moderate crisis–plus
2016:H1 Minor crisis–plus
2016:H2 Minor crisis–reg.
2017:H1 Minor crisis–minus
2017:H2 Minor crisis–minus

Hungary
2008:H2 Minor crisis–reg.
2009:H1 Moderate crisis–reg.
2009:H2 Minor crisis–minus
2010:H1 Minor crisis–reg.
2010:H2 Credit disrupt.–plus
2011:H1 Credit disrupt.–plus
2011:H2 Minor crisis–plus
2012:H1 Minor crisis–plus
2012:H2 Minor crisis–reg.
2013:H1 Minor crisis–plus
2013:H2 Minor crisis–plus
2014:H1 Minor crisis–plus
2014:H2 Minor crisis–plus
2015:H1 Minor crisis–reg.
2015:H2 Credit disrupt.–plus
2016:H1 Credit disrupt.–minus

Iceland
2013:H1 Credit disrupt.–minus

Ireland
2013:H1 Minor crisis–plus
2013:H2 Minor crisis–plus
2014:H1 Minor crisis–minus
2014:H2 Credit disrupt.–plus
2015:H1 Credit disrupt.–plus
2015:H2 Credit disrupt.–minus
2016:H1 Credit disrupt.–plus
2016:H2 Credit disrupt.–reg.
2017:H1 Credit disrupt.–plus
2017:H2 Credit disrupt.–plus

Italy
2013:H1 Moderate crisis–reg.
2013:H2 Minor crisis–reg.
2014:H1 Minor crisis–minus
2014:H2 Minor crisis–minus
2015:H1 Minor crisis–minus
2015:H2 Minor crisis–minus
2016:H1 Minor crisis–minus
2016:H2 Minor crisis–minus
2017:H1 Credit disrupt.–plus
2017:H2 Credit disrupt.–reg.

Korea, South
1997:H1 Credit disrupt.–plus
1997:H2 Moderate crisis–reg.
1998:H1 Major crisis–minus
1998:H2 Moderate crisis–plus
1999:H1 Moderate crisis–reg.
1999:H2 Moderate crisis–minus
2000:H1 Credit disrupt.–plus
2000:H2 Minor crisis–reg.
2001:H1 Minor crisis–plus
2001:H2 Minor crisis–minus
2002:H1 Credit disrupt.–plus
2003:H2 Minor crisis–minus
2004:H1 Minor crisis–plus
2004:H2 Credit disrupt.–plus
2005:H1 Credit disrupt.–minus
2008:H2 Minor crisis–plus
2009:H1 Minor crisis–minus
2009:H2 Credit disrupt.–minus
2012:H2 Credit disrupt.–minus

Table 1. Financial Distress in Countries Belonging to the Organization for Economic Cooperation and Development, 2013:H1–2017:H2 (and in the Added Countries Starting When Information Is Available)[a] (*Continued*)

Mexico
1995:H2 Minor crisis–plus
1996:H1 Moderate crisis–minus
1996:H2 Minor crisis–reg.
1997:H1 Credit disrupt.–reg.
1997:H2 Minor crisis–minus
1998:H1 Credit disrupt.–reg.
2008:H2 Credit disrupt.–minus

Netherlands
2013:H1 Minor crisis–minus
2013:H2 Minor crisis–reg.
2014:H1 Credit disrupt.–minus
2014:H2 Minor crisis–minus
2015:H1 Credit disrupt.–reg.
2015:H2 Credit disrupt.–minus

Poland
2008:H2 Credit disrupt.–reg.
2009:H1 Minor crisis–reg.
2009:H2 Credit disrupt.–reg.
2011:H2 Credit disrupt.–reg.
2012:H1 Credit disrupt.–minus
2012:H2 Credit disrupt.–minus
2013:H2 Credit disrupt.–minus

Portugal
2013:H1 Minor crisis–plus
2013:H2 Minor crisis–reg.
2014:H1 Minor crisis–minus
2014:H2 Minor crisis–minus
2015:H1 Credit disrupt.–plus
2015:H2 Credit disrupt.–plus
2016:H1 Minor crisis–minus
2016:H2 Minor crisis–minus
2017:H1 Minor crisis–minus
2017:H2 Credit disrupt.–plus

Slovak Republic
2008:H2 Minor crisis–minus
2009:H1 Credit disrupt.–minus
2012:H1 Minor crisis–minus
2012:H2 Credit disrupt.–minus
2014:H1 Credit disrupt.–minus

Spain
2013:H1 Minor crisis–plus
2013:H2 Moderate crisis–minus
2014:H1 Minor crisis–reg.
2014:H2 Credit disrupt.–plus
2015:H1 Credit disrupt.–minus

United Kingdom
2013:H1 Credit disrupt.–minus
2013:H2 Credit disrupt.–minus
2014:H1 Credit disrupt.–minus

Sources: *OECD Economic Outlook*; authors' analysis.

a. H = half year. This table shows the nonzero values for our scaled measure of financial distress for 30 OECD countries from 2013:H1 to 2017:H2. It also shows all nonzero values for the 6 countries added to the sample going back to when they enter the sample (2003:H2 for the Czech Republic, 1998:H1 for Hungary, 1994:H1 for Mexico, 1998:H1 for Poland, 2003:H2 for the Slovak Republic, and 1996:H2 for South Korea). See the text and online appendix A for details of the derivation of the measure. For the values of the measure of financial distress for the original 24 countries in our sample for 1967:H1 to 2012:H2, see Romer and Romer (2017, table 1, 3081–82).

nonzero values for the 6 added countries for all years that information is available. Online appendix A contains our reasoning for all the observations added to the sample that we classify as having a positive level of financial distress.[3] The inclusion of 5 more years and 6 additional countries increases the number of observations covered by our measure by

3. The online appendixes for this and all other papers in this volume may be found at the *Brookings Papers* web page, www.brookings.edu/bpea, under "Past BPEA Editions."

21 percent. However, because there was almost no financial distress in the early part of our sample, the amount of distress covered by the measure increases by much more: the number of observations where our measure is strictly positive rises by 50 percent.

Figure 1 shows the expanded measure of financial distress for the 30 countries for 1980–2017, which is the period that we focus on in this paper. The top panel shows the measure from the start of the period through 2005, when financial distress never affected more than a few countries simultaneously. The bottom panel shows the series for 2006 through the end of the sample, when every country in our sample experienced at least some distress. Relative to our previous sample, there are now two additional episodes of high distress in the 1990s, one in Mexico and one in South Korea. Expanding the sample of countries and going through 2017 also provides a more complete picture of the global financial crisis. The bottom panel of the figure shows that there is tremendous variation in how quickly financial distress faded after 2008. Some countries where the crisis was initially very severe, such as the United States and the United Kingdom, became largely free of distress within a few years. Conversely, Greece, Ireland, Italy, and Portugal still had some financial distress at the end of 2017. Furthermore, though all the added countries experienced some distress after 2008, only Hungary experienced distress of 7 or above on our scale (a lower-level moderate crisis).

I.B. The Average Aftermath of a Financial Crisis

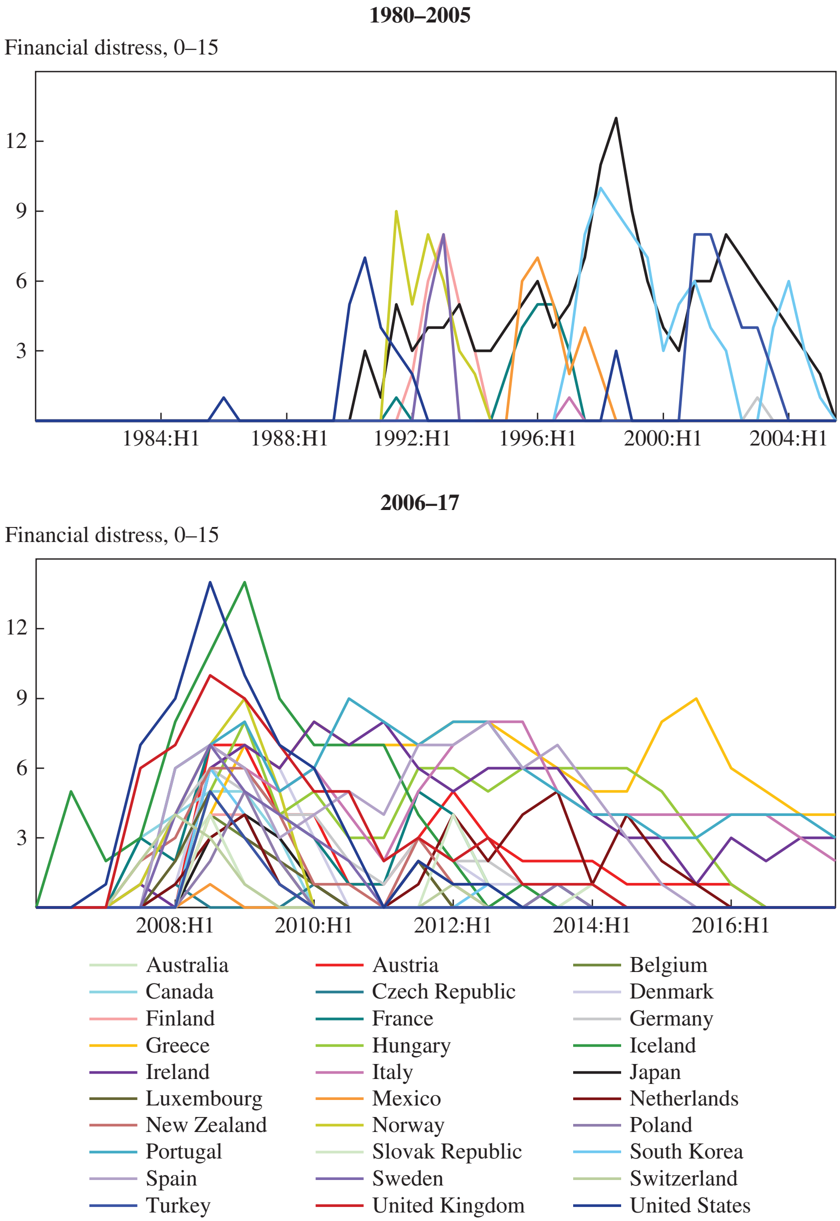
Because we have expanded the sample substantially, a useful first step is to see if using the new sample alters our original findings on the average aftermath of a financial crisis. To investigate the average aftermath, we estimate the following Jordà local projection panel regression:

$$y_{j,t+h} = \alpha_j^h + \gamma_t^h + \beta^h F_{j,t} + \sum_{k=1}^{4} \varphi_k^h F_{j,t-k} + \sum_{k=1}^{4} \theta_k^h y_{j,t-k} + e_{j,t}^h, \tag{1}$$

where the j subscripts index countries, the t subscripts index time, and the h subscripts and superscripts denote the horizon (half years after time t). The term $y_{j,t+\mathrm{h}}$ is the logarithm of real GDP in country j at time $t + h$. The term $F_{j,t}$ is the financial distress variable for country j at time t. The αs are country fixed effects, and the γs are time fixed effects. We include four lags of both output and distress to account for the usual dynamics of these series.

We estimate equation 1 separately for horizons 0 to 10 (that is, up through five years after time t). The sequence of β^hs from these 11 regressions

Figure 1. Measure of Financial Distress for an Extended Sample of Countries and an Extended Time Period[a]



Sources: *OECD Economic Outlook;* authors' analysis.

a. H = half year. This figure shows semiannual values for the new scaled measure of financial distress for 30 countries that belong to the Organization for Economic Cooperation and Development. See the text and online appendix A for the details of the derivation of the measure.

provides a nonparametric estimate of the impulse response function of output to an innovation in financial distress of one step. To get a sense of the aftermath of a “crisis,” we multiply the point estimates by 7, which is the number on our scale corresponding to the start of the “moderate crisis” category. Importantly, the specification includes as part of the average aftermath of distress any contemporaneous relationship between output and distress. Because distress is almost surely at least somewhat endogenous, the estimated impulse response function should thus be viewed as an upper bound of any causal effect of distress on economic activity.[4] The GDP data are from the OECD.[5] For consistency with our subsequent empirical work, which uses fiscal data that only begin in 1980, we restrict all the data used in the estimation to the period 1980–2017.

The top panel of figure 2 shows the estimated impulse response function (along with the 2-standard-error confidence bands) using our full set of 30 OECD countries. This panel also shows the results for our original sample of 24 countries. For the full sample of countries, the aftermath of a realization of financial distress of a 7 on our scale is a substantial and persistent decline in real GDP. The peak fall in output after a crisis is a decline of just over 4 percent, and is highly significant ($t = -4.1$).[6]

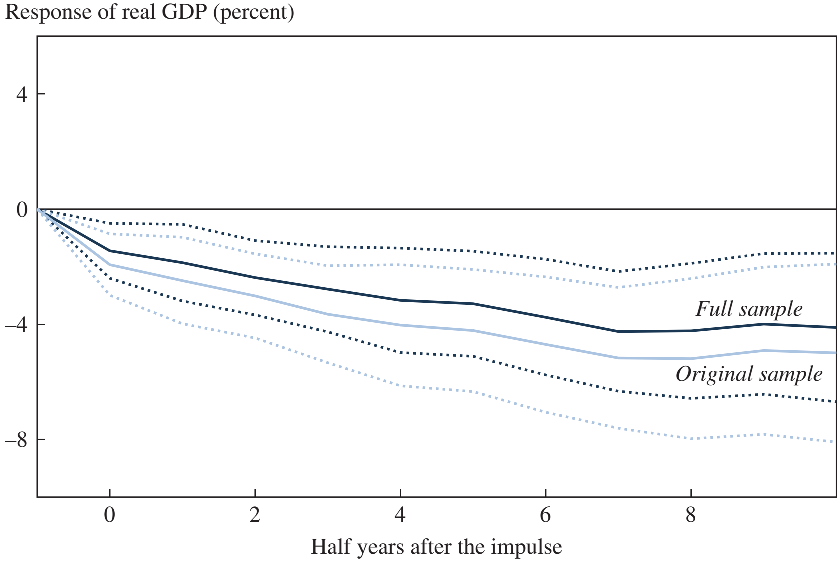
4. Romer and Romer (2017) provide an extensive discussion of causation and timing. We find that excluding the contemporaneous relationship between output and financial distress reduces the negative aftermath of a crisis by nearly half. This suggests that endogeneity issues are indeed important, and that the true causal impact of financial distress is substantially smaller than the aftermath as estimated in equation 1. Unfortunately, our narrative source is not adequate for identifying genuinely exogenous episodes of financial distress or determining if such episodes even exist in the postwar period.

5. See https://stats.oecd.org. The data are from the Quarterly National Accounts Dataset, series VPVOBARSA. GDP data are missing for a few countries in certain years. Because the financial distress variable is semiannual (roughly corresponding to June and December), we convert the GDP data to semiannual as well (using the observations for the second and fourth quarters of each year). Ireland’s GDP jumped more than 20 percent in 2015:Q1, due largely to the relocation of many companies’ intellectual property to Ireland. Because this is such an extreme observation and is unrelated to the normal determinants of output movements, we do not use Irish data after 2014:Q4.

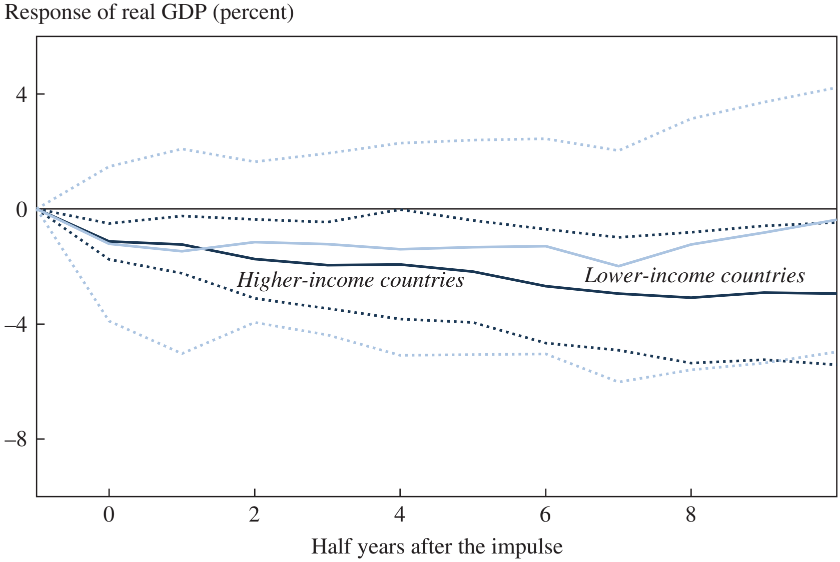
6. Throughout, we report results based on heteroskedasticity-corrected standard errors, which are generally considerably larger than the conventional standard errors from our regressions. We have also examined various ways of correcting the standard errors for serial correlation (Newey–West and Hansen–Hodrick standard errors and clustering by country), as well as clustering by time period. However, because of the inclusion of lags in our regressions, we are focusing on responses to innovations in our variables (in this case, financial distress), which are by construction roughly serially uncorrelated. As a result, one would not expect serial correlation of the residuals to cause important bias in the standard errors. And indeed, the various alternatives do not change the standard errors systematically relative to the heteroskedasticity-corrected ones.

Figure 2. The Behavior of Real GDP after a Financial Crisis[a]

The full sample (30 countries) and the original sample (24 countries)



Samples of countries with higher and lower incomes than Greece



Source: Authors' calculations.

a. This figure shows the impulse response function of real GDP to an impulse of 7 in financial distress based on estimation of equation 1 over the period 1980:H1–2017:H2 for different samples of countries; H = half year. The dotted lines show the 2-standard-error confidence bands.

This estimated aftermath is noticeably less severe than we found in our 2017 paper, which was a decline of 6.0 percent.[7] There are three changes in the estimation relative to our previous paper: a larger sample of countries; a different time period (the original period was 1967–2012); and revisions to the GDP data. The top panel of figure 2 shows that considering only the original sample of 24 countries (but for the 1980–2017 period) results in a decline in GDP after a crisis of 5.2 percent ($t = -3.7$). Thus, the new sample of countries is an important source of the difference between the new estimates and the original ones.

Because the added countries are at the lower end of the spectrum of per capita GDP, it is useful to consider where there are systematic differences in the aftermath of financial crises between higher-income and lower-income countries. Because Greece is an influential observation in whatever sample it is in, it is natural to use it as the dividing line between higher-income and lower-income countries, and to leave it out of both samples. To classify countries, we therefore compare their per capita GDP in 1992 (the first year for which there are annual data on GDP per capita for all 30 countries) to that of Greece.[8] Eight countries had a lower GDP per capita than Greece: the Czech Republic, Hungary, Mexico, Poland, Portugal, the Slovak Republic, South Korea, and Turkey.

The bottom panel of figure 2 shows the estimated impulse response functions for higher-income and lower-income countries. Both types of countries have a smaller negative aftermath of a crisis than the full sample, consistent with the notion that Greece's terrible downturn after the global financial crisis pulls down the average aftermath in the full sample noticeably.[9] The point estimates for the two types of countries, however, are quite different. The aftermath of a crisis is more negative and more persistent in higher-income countries than in lower-income ones. Indeed,

7. Consistent with what we found earlier (Romer and Romer 2017), if we exclude the contemporaneous relationship between output and financial distress in the estimated aftermath, the average aftermath of a crisis is substantially less severe than the baseline estimates. Using the expanded sample considered in this paper, the peak fall in output using this specification is just over 2 percent ($t = -1.8$).

8. We use GDP per head (current dollars) from the OECD (https://stats.oecd.org).

9. Although it does not make sense to ignore the evidence from Greece's experience following the global financial crisis entirely, the fall in its output was so extreme that it is natural to wonder if Greece could be driving our results. We have therefore reestimated all our key equations excluding Greece from the sample. The general pattern is that dropping Greece weakens the results somewhat, but does not change them qualitatively. Perhaps the most interesting exception is that in some of the regressions in section III, excluding Greece actually slightly strengthens the relationship between direct measures of market access and the fiscal policy response to financial distress.

for lower-income countries, the negative aftermath is completely undone within five years of the crisis, whereas for higher-income countries it is not undone at all. Not surprisingly, given the smaller sample, the 2-standard-error bands are very wide for the lower-income country sample. Nevertheless, the finding that the negative aftermath of a crisis appears milder in less advanced countries goes against the common view that crises are more devastating in developing economies.

I.C. Variation in Aftermaths

The variation in the aftermaths of crises between higher-income and lower-income countries is consistent with the finding in our 2017 paper that there is, in general, substantial variation in aftermaths across crisis episodes. One way to show this variation is to focus on the 22 episodes of high financial distress (which we define as a reading of 7 or greater on our scale of 0–15) in our sample. We consider forecasts of real GDP in each episode based on the estimates from equation 1. In forming the forecasts, we use the realization of the distress variable up through the half year that it reaches 7 or higher, and actual GDP up through one half year before this occurs. We then calculate forecast residuals as actual GDP minus the forecast, so negative residuals correspond to actual GDP being lower than the forecast.

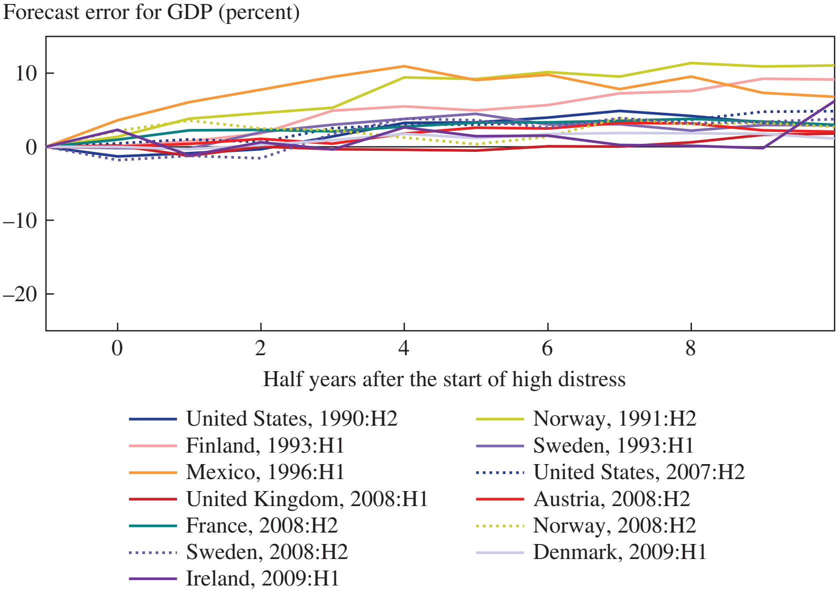
Because we include actual distress up through the start of the forecast, the forecasts take into account that these are all crisis episodes. As a result, the forecast errors are roughly mean zero across episodes. Nevertheless, there is substantial variation in the errors across the episodes.[10] This variation is the result of differences both in how financial distress itself evolves in each episode, and in how GDP responds to a given level of distress.

Figure 3 shows the forecast errors in the various episodes. We divide them into the cases with very small negative or positive forecast errors and those with substantial negative forecast errors. Even within these two groups, there is a wide range of outcomes. Among the episodes of relatively small or positive forecast errors shown in the top panel of the figure, there are cases like Sweden after its 1993 crisis, where the forecast errors are small and negative in the immediate aftermath, but small and positive thereafter. Conversely, Mexico (after its 1996 crisis), Norway

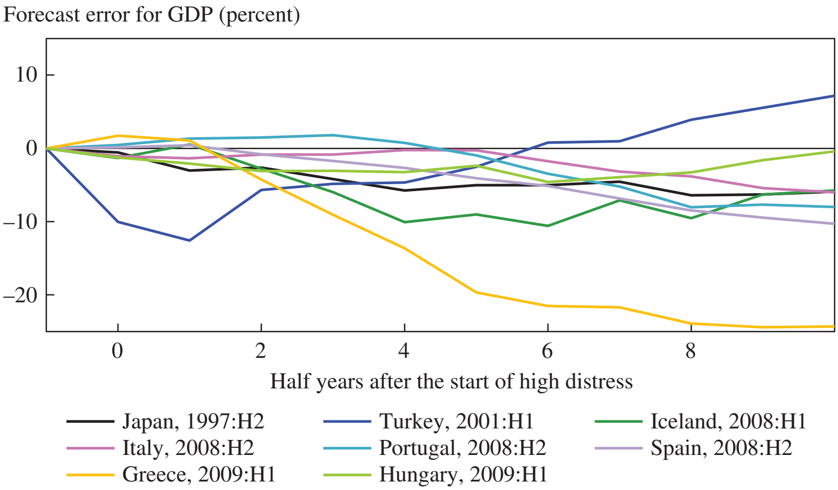
10. In this exercise, South Korea is excluded. Its crisis in 1997 occurred just a year after it joined the OECD. As a result, it lacks the four lags of the distress variable needed to construct the forecast.

Figure 3. GDP Forecast Errors for Episodes of High Financial Distress[a]

Cases with positive or small negative forecast errors



Cases with substantial negative forecast errors



Source: Authors' calculations.

a. H = half year. This figure shows the forecast errors for real GDP based on estimation of equation 1 over the period 1980:H1–2017:H2. Forecast errors are shown for the 21 episodes where distress reached 7 or above and sufficient lags are available to construct the forecasts. The forecasts use the realization of the distress variable up through the half year that it reached 7 or higher, and actual GDP up through one half year before that occurs. The dates given are when distress first reached 7 or above.

(after its 1991 crisis), and Finland (after its 1993 crisis) all experienced actual growth much higher than the forecast during almost all five years after the start of high distress.

There is even greater variation in aftermaths among the episodes of substantial negative forecast errors shown in the bottom panel of figure 3. After its crisis in 2009:H1, Greece experienced GDP declines far worse and more persistent than those predicted using equation 1. Likewise, Spain, Portugal, and Italy (after the global financial crisis) and Japan (after its 1997 crisis) experienced severe and persistent negative forecast errors. Two of the lower-income countries in this group (Turkey after its 2001 crisis and Hungary after its 2009 crisis) show another interesting pattern. There is a short-run drop in output greater than the forecast (in the case of Turkey, dramatically greater), but then substantial recovery. Indeed, after its catastrophic initial decline, Turkey experienced growth almost equally dramatically above the forecast.

II. The Importance of Fiscal Space

The evidence in the preceding section shows that although the aftermath of financial crises is in general quite negative, there is tremendous variation in the severity and persistence of the output declines after high financial distress. We turn now to the role that fiscal space plays in explaining this variation. The analysis in this section largely extends some of the findings of our 2018 paper using our larger number of countries and longer time period. Sections III and IV consider the issue of why space matters.

II.A. Definition of Fiscal Space

We think of fiscal space as the room a country has to use fiscal policy to stimulate the economy or to undertake a bailout and recapitalization of its financial sector. For our analysis, we define fiscal space as the negative of the ratio of gross government debt to GDP. Thus, it is a continuous measure, with fiscal space declining linearly with the debt-to-GDP ratio.

There are obviously many other ways to define fiscal space. For example, in our 2018 paper, we consider using net debt in place of gross debt, and investigate replacing the linear specification with more complicated threshold-type formulations.[11] Later in this section, we consider whether the prior budget surplus might be an added component of fiscal space.

11. We find that these variations have little effect on our estimates of the role of fiscal space, and so do not repeat them in this study.

And, in section III, we explore whether more direct indicators of sovereign market access dominate the gross debt-to-GDP ratio in determining the postcrisis behavior of fiscal policy. But a country's gross debt load is a fundamental and intuitive way to conceptualize fiscal space.

A virtue of using the debt-to-GDP ratio (or, more precisely, the negative of the debt-to-GDP ratio) as the measure of fiscal space is that this ratio is determined in large part by past policy decisions and more long-run features of a country's policymaking process. It captures the fact that some countries (like Greece and Italy) perennially run deficits, while others (like South Korea and Germany) typically pursue balanced budgets. It obviously also responds somewhat to movements in output and fiscal policy during and after financial crises, but it is typically slower-moving and less cyclically sensitive than such indicators as the budget surplus and interest rates. To further strengthen the exogeneity of the debt-to-GDP ratio to policy decisions made in the context of crises, in the regressions below we always use the ratio at the end of the previous calendar year.[12]

Fiscal data are generally not available on a comparable basis for a wide range of countries before 1980. As a result, our analysis focuses on the period 1980–2017. Data on gross government debt for our sample of countries for most of the period starting in 1980 are available from the IMF's World Economic Outlook database.[13] When values going all the way back to 1980 are not available from the IMF, we extend the series back using data from the OECD when possible.[14] The resulting debt series covers 95 percent of the observations since 1980 for which our measure of financial distress is available.

II.B. Fiscal Space and the Response of GDP to Financial Distress

To see if fiscal space explains some of the variation in the aftermaths of financial crises, we augment equation 1 to include an interaction term

12. The reason for using the debt-to-GDP ratio at the end of the previous year is that the debt-to-GDP ratio data are annual, end-of-year values. Thus we use the debt ratio at the end of period $t-1$ when period t corresponds to the first half of the year, and the ratio at the end of period $t-2$ when period t corresponds to the second half of the year.

13. We use the data from the October 2018 edition of the database, https://www.imf.org/external/pubs/ft/weo/2018/02/weodata/index.aspx.

14. The OECD data are available from https://stats.oecd.org/. For a few countries, gross debt data for early years of the sample are available in earlier published editions of the *OECD Economic Outlook*, but not from the OECD website. In such cases, we use those data (specifically, data from the December 2002 and December 1996 editions of the *Economic Outlook*). We join the series using splices in levels, working backward in time through the various sources.

between financial distress and the negative of the debt-to-GDP ratio at the end of the previous year. The coefficients on this interaction term measure how the response of output to distress varies with fiscal space. In addition to the interaction term, we also include the negative of the debt ratio alone, again as of the end of the previous year. Thus, we estimate:

$$(2) \qquad y_{j,t+h} = \alpha_j^h + \gamma_t^h + \vartheta^h S_{j,t} + \beta^h F_{j,t} + \delta^h \left(F_{j,t} \cdot S_{j,t} \right) + \sum\nolimits_{k=1}^{4} \rho_k^h S_{j,t-k}$$
$$+ \sum\nolimits_{k=1}^{4} \varphi_k^h F_{j,t-k} + \sum\nolimits_{k=1}^{4} \omega_k^h \left(F_{j,t-k} \cdot S_{j,t-k} \right) + \sum\nolimits_{k=1}^{4} \theta_k^h y_{j,t-k} + e_{j,t}^h,$$

where $S_{j,t}$ is our measure of fiscal space in country j in half year t, and all other variables are as before. We again estimate the relationship for the horizons $h = 0$–10.

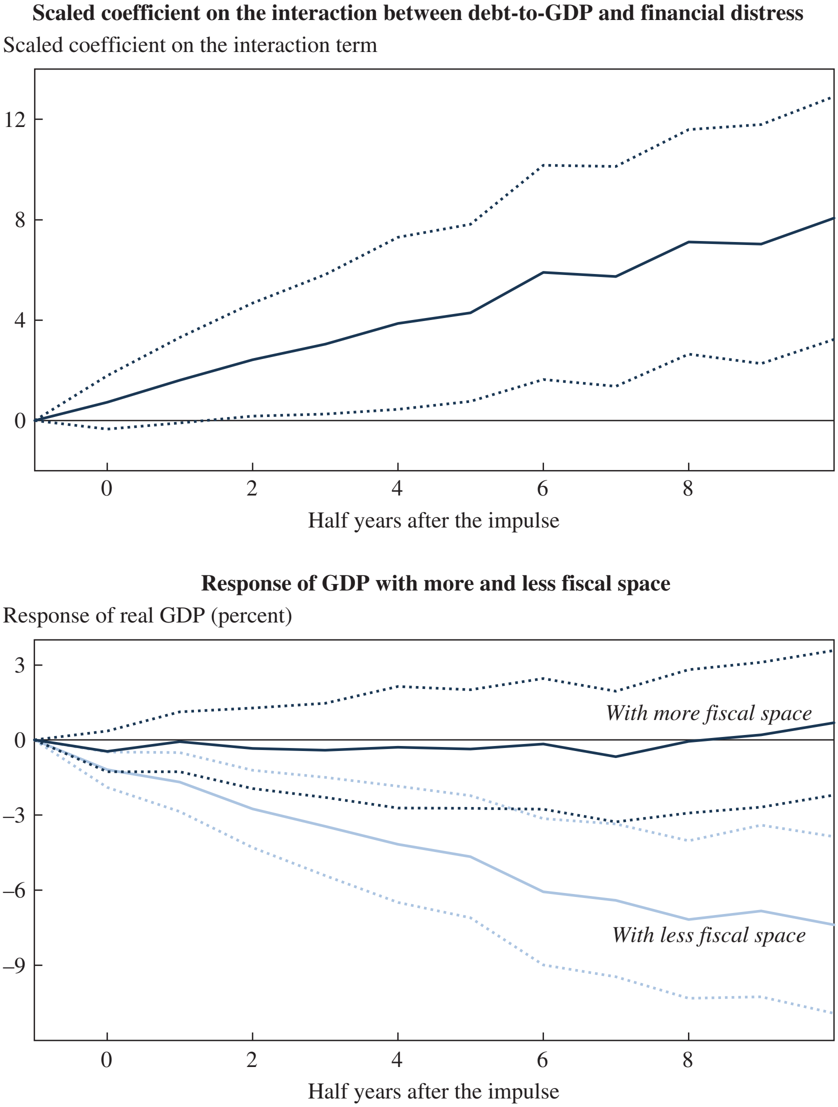
The top panel of figure 4 shows the estimated coefficient on the interaction term (d^h) at the various horizons, together with the 2-standard-error bands. To make it easier to interpret the coefficients, we multiply them by a realization of the interaction term of twice the standard deviation of the gross debt-to-GDP ratio in our sample (which is roughly 35 percentage points), times 7. The factor of 7 accounts for the fact that we are interested in the impact of a fairly substantial rise in financial distress. Thus, the reported numbers can be interpreted as how the behavior of output after a financial crisis varies with a 2-standard-deviation increase in fiscal space.

Figure 4 shows that the scaled coefficient on the interaction term is positive at all horizons and statistically significant after horizon 1 (with a maximum t statistic over 3). The fact that the coefficients are positive means that the fall in GDP after a crisis is smaller when the negative of the debt-to-GDP ratio is less negative—that is, when there is more fiscal space.

The bottom panel of figure 4 presents another way of visualizing the implications of the estimates for the importance of fiscal space. It shows the impulse response function of GDP based on equation 2 to an innovation in financial distress of 7, including both the direct effect of distress (the β^hs) and the interaction effect (the d^hs) for two cases: when the debt-to-GDP ratio is 1 standard deviation above the sample mean ("less fiscal space"), and when it is 1 standard deviation below the sample mean ("more fiscal space"). These correspond to debt ratios of roughly 25 and 95 percent.

The bottom panel of figure 4 shows that the aftermath of a financial crisis is dramatically different in the two cases. GDP typically falls about 7 percent after a realization of 7 on our scale of financial distress when the debt-to-GDP ratio is 1 standard deviation above the mean, but by less

Figure 4. The Relationship between Real GDP after a Financial Crisis and Fiscal Space[a]



Source: Authors' calculations.

a. This figure is based on estimates of equation 2 over the period 1980:H1–2017:H2; H = half year. The top panel shows the estimated values of δ^h for different values of h, scaled by 7 times twice the sample standard deviation of the gross debt-to-GDP ratio. The bottom panel shows the implied impulse response functions of real GDP to an impulse of 7 in financial distress for a country with a debt-to-GDP ratio in the previous calendar year 1 standard deviation below the sample mean ("with more fiscal space"), and for a country with a debt ratio 1 standard deviation above the sample mean ("with less fiscal space"). The dotted lines show the 2-standard-error confidence bands.

than 1 percent when the debt ratio is 1 standard deviation below the mean. Although these two cases represent a sizable difference in the debt ratio, the difference is by no means extreme. And because space is assumed to decline linearly with the debt-to-GDP ratio, a smaller or larger difference would imply a proportionally smaller or larger estimated difference in the aftermath of a crisis.

The two cases presented in the bottom panel of figure 4 explain the logic in the construction of the figure's top panel of multiplying the estimated coefficient on the interaction term by twice the sample standard deviation of the debt ratio, and then by 7. By doing this, we show precisely the difference in the impulse response functions of output to a financial crisis (defined as an innovation of 7 in our measure) between the cases of more and less fiscal space. That is, the top panel shows the difference between the two impulse response functions presented in the bottom panel, together with the 2-standard-error bands.

II.C. Fiscal Space and the Response of Fiscal Policy to Financial Distress

The most obvious mechanism by which fiscal space could affect the aftermath of a crisis is by enabling or limiting fiscal stimulus and financial rescue. It is therefore natural to examine how the behavior of fiscal policy after a crisis varies with fiscal space.

To do this, we run interaction regressions like those for GDP, but using a measure of the change in the high-employment surplus as the dependent variable. Official estimates of the high-employment surplus are available on a consistent basis for a large number of countries in our sample only for relatively recent years. For this reason, we consider an approximation. For each horizon h that we consider, we use as the left-hand-side variable the change in the actual budget surplus (as a share of GDP) from $t - 1$ to $t + h$, minus the percentage change in real GDP times an estimate of the cyclical sensitivity of the surplus to GDP. That is, we estimate

$$
\begin{aligned}
(3) \quad (B_{j,t+h} - B_{j,t-1}) - \tau \cdot (y_{j,t+h} - y_{j,t-1}) = {} & \alpha_j^h + \gamma_t^h + \vartheta^h S_{j,t} + \beta^h F_{j,t} + \delta^h (F_{j,t} \cdot S_{j,t}) \\
& + \sum_{k=1}^{4} \rho_k^h S_{j,t-k} + \sum_{k=1}^{4} \varphi_k^h F_{j,t-k} \\
& + \sum_{k=1}^{4} \omega_k^h (F_{j,t-k} \cdot S_{j,t-k}) \\
& + \sum_{k=1}^{4} \theta_k^h (\Delta B_{j,t-k} - \tau \cdot \Delta y_{j,t-k}) + e_{j,t}^h ,
\end{aligned}
$$

where $B_{j,t}$ is the budget surplus as a share of GDP in country j in period t, and t is the assumed sensitivity of the budget surplus to real activity. We estimate equation 3 for horizons $h = 0–10$.

This specification omits the growth of potential output. That is, it leaves out a $\tau \cdot (\overline{y}_{j,t+h} - \overline{y}_{j,t-1})$ term in the calculation of the change in the high-employment surplus, where $\overline{y}$ is potential output. If trend growth in each country is constant over our sample period, however, this term will be captured by the country fixed effects (the α_j^hs). Thus, this method of estimating the change in the high-employment surplus makes sense as long as trend or potential growth for each country does not change greatly over our sample period. Based on the evidence found by Nathalie Girouard and Christophe André (2005), a reasonable estimate of t for OECD countries is 0.4. Finally, note that because we consider the change in the high-employment surplus over progressively longer horizons, the estimates from equation 3 inherently show how the cumulative response of the high-employment surplus depends on financial distress and its interaction with fiscal space.[15]

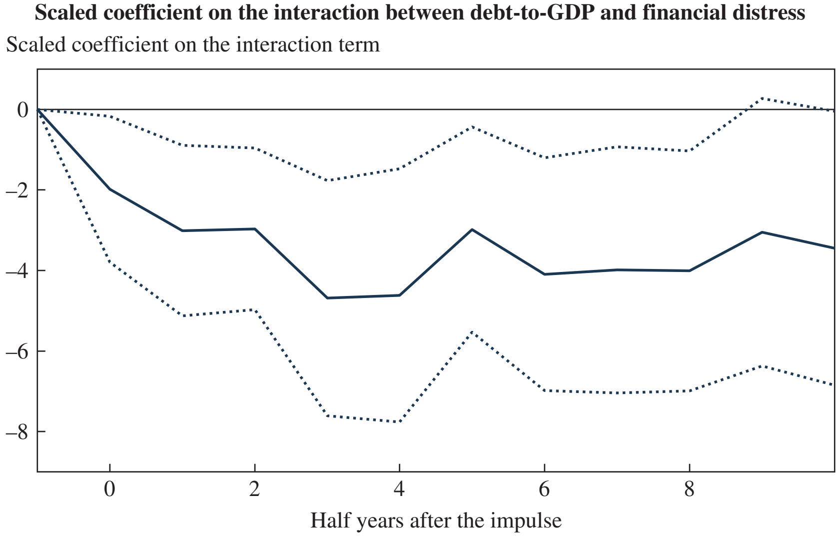
We obtain data on the budget surplus from the same sources as our data on the debt-to-GDP ratio. Specifically, the data are from the IMF's World Economic Outlook database when available, supplemented with data from the OECD when those go back further.[16] The resulting series covers 97 percent of the observations since 1980 for which our measure of financial distress is available.

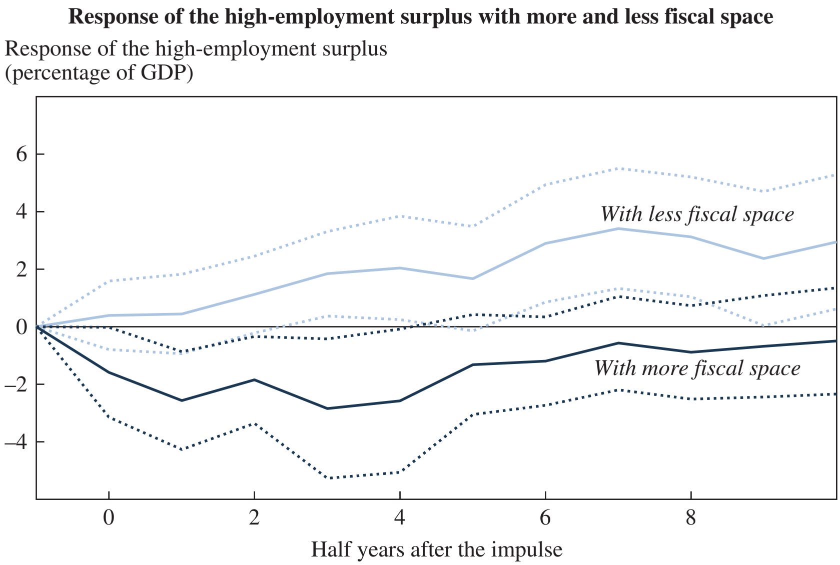
The top panel of figure 5 shows the estimates of the d^hs, the coefficients on the interaction term. We again multiply the estimates by 7 and by 2 times the standard deviation of the debt-to-GDP ratio to aid interpretation. The estimates are negative and highly statistically significant. The fact that the estimates are negative means that countries with lower debt ratios (and thus with more fiscal space) respond to financial distress with

15. We also examine the effects of allowing t to vary across countries using the estimates from Girouard and André (2005). Because Girouard and André do not report ts for Mexico and Turkey, we are forced to drop these two countries from our sample. In all cases, the results are extremely similar to our baseline ones for the same sample, although they are typically very slightly stronger.

16. All data were downloaded November 11, 2018, and the data from earlier published versions of the *OECD Economic Outlook* are again from the December 2002 and December 1996 editions. For the IMF data, we use the series "General Government Net Lending/Borrowing," and for the *Economic Outlook*, we use the series "Financial Balance." We again join the various series using splices in levels, working backward in time through the sources. One small difference from our series for the debt-to-GDP ratio is that we do not use any current OECD data from https://stats.oecd.org/.

Figure 5. The Relationship between the High-Employment Surplus after a Financial Crisis and Fiscal Space[a]





Source: Authors' calculations.

a. This figure is based on estimates of equation 3 over the period 1980:H1–2017:H2; H = half year. The top panel shows the estimated values of δ^h for different values of h, scaled by 7 times twice the sample standard deviation of the gross debt-to-GDP ratio. The bottom panel shows the implied impulse response functions of the high-employment surplus to an impulse of 7 in financial distress for a country with a debt-to-GDP ratio in the previous calendar year 1 standard deviation below the sample mean ("with more fiscal space"), and for a country with a debt ratio 1 standard deviation above the sample mean ("with less fiscal space"). The dotted lines show the 2-standard-error confidence bands.

lower (or more negative) high-employment surpluses. That is, they run more expansionary fiscal policy.

The bottom panel of figure 5 shows the implications of the estimates for the behavior of the high-employment surplus after an innovation of 7 in the new measure of financial distress, including both the direct impact of distress and the interaction term. We again consider the cases where the debt-to-GDP ratio is 1 standard deviation below its mean ("more fiscal space") and where it is 1 standard deviation above the mean ("less fiscal space"). The figure shows just how important the interaction with fiscal space is. A country facing high financial distress with a debt-to-GDP ratio 1 standard deviation below the mean cuts its high-employment surplus by 2 to 3 percent of GDP; a country facing high distress with a debt ratio 1 standard deviation above the mean runs contractionary fiscal policy, with its high-employment surplus rising by over 3 percent of GDP.

Given that both GDP and the high-employment surplus after crises vary strongly with the prior debt-to-GDP ratio, it is natural to think that there is a link between the two. A large body of literature finds that changes in taxes and government spending have powerful effects on real output (for example, Fisher and Peters 2010; Romer and Romer 2010; Ramey 2011; and Guajardo, Leigh, and Pescatori 2014).[17] Thus, it is highly likely that output declines after crises are smaller when a country faces a crisis with low debt *because* low-debt countries use fiscal policy aggressively to mitigate the impact of the crisis and rescue the financial system, while high-debt countries pursue contractionary fiscal policy.

II.D. The Role of the Prior Budget Surplus

Another variable that may affect a country's ability or willingness to use expansionary fiscal policy in response to financial distress is the level of its budget surplus before the distress occurs. For a given degree of fiscal expansion, the resulting deficit will be larger when the prior surplus is smaller. To the extent that a larger deficit increases difficulties with market access or makes policymakers want to pursue less expansionary

17. Another large body of literature uses cross-sectional data to investigate the impact of changes in government spending on output and employment (Nakamura and Steinsson 2014; Chodorow-Reich and others 2012; Suárez Serrato and Wingender 2016). These cross-sectional studies typically find a fiscal multiplier of about 1.5. Chodorow-Reich (2019) argues that the cross-sectional multiplier is an approximate lower bound on the aggregate multiplier for cases where monetary policy does not respond to fiscal policy (which applies to many of the crises in our sample).

policy, a smaller prior surplus could therefore lead to a less expansionary response to distress.

To investigate this issue, we estimate variants of equation 3 using the surplus as a share of GDP in the previous year in place of, or in addition to, the negative of the previous year's debt-to-GDP ratio.[18] The results show a strong relationship between the prior surplus and the fiscal policy response to financial distress. When we use the prior surplus in place of the negative debt ratio, it is highly significant and quantitatively important. The *t* statistic for the coefficient on the interaction term between distress and the previous year's surplus ranges from 2.5 to 3.7 (with the exception of horizon 0, when it is 1.8), and the point estimates indicate that an improvement of 2 standard deviations in the prior surplus (roughly 9 percentage points) is associated with a more expansionary response of the high-employment surplus to an innovation of 7 in distress that is smaller than what we find for a 2-standard-deviation improvement in the prior debt ratio, but still large—2 to 3 percent of GDP.

When we include both measures, the point estimates on both are quantitatively large, and both are statistically significant. The point estimates suggest that the prior debt ratio is moderately more important quantitatively than the prior surplus; however, the prior surplus is somewhat more statistically significant. The null hypothesis that neither variable is related to the fiscal response to distress is overwhelmingly rejected, with *p* values less than 0.001 at most horizons. Thus, bringing the prior surplus into the analysis strengthens the finding that there is a powerful relationship between a country's fiscal situation and its fiscal response to financial distress.

At the same time, we are reluctant to place too much weight on the findings involving the prior budget surplus. As discussed above, the debt ratio is determined largely by long-term forces. The prior surplus, in contrast, is heavily influenced by recent policy decisions. One concrete concern is that if policymakers have information about current or prospective financial distress before the distress is reflected in our measure, they may pursue fiscal expansion, and thus run large deficits, before our measure of distress rises. If so, the finding that a smaller prior surplus is associated with a less expansionary response to distress could reflect not a causal impact of the

18. When we use the prior surplus in place of the prior debt-to-GDP ratio, we replace the negative of the debt ratio in the prior year (S) with the surplus-to-GDP ratio in the prior year whenever it appears in equation 3. When we use it in addition to the debt ratio, we add the corresponding variable using the prior surplus-to-GDP ratio whenever a variable using the prior debt-to-GDP ratio appears in equation 3.

prior surplus, but merely the fact that countries that act before our measure of distress rises pursue less additional expansion when the increase in our measure occurs. Because of the potential difficulties with interpretations of correlations involving the prior surplus, in the remainder of the paper we continue to focus on just the prior debt ratio.[19]

II.E. Looking at Episodes of High Distress

One way to get a sense of the sources of the baseline fiscal space regression results and to gain more confidence that they reflect genuine patterns in the data is to look at the behavior of debt, the high-employment surplus, and financial distress in the episodes of high distress in our sample. Specifically, we look at the 22 cases where distress reached 7 or more.[20]

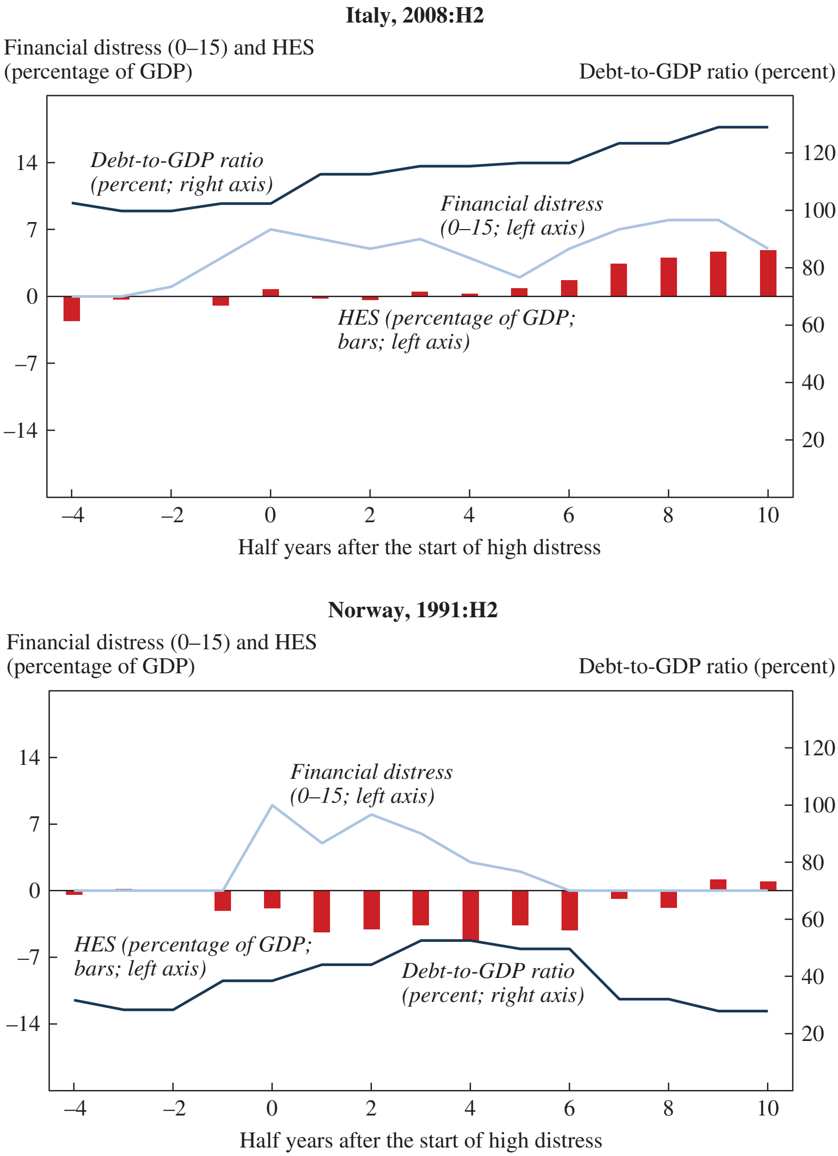
Figure 6 shows two cases where the overall patterns fit straightforwardly with the regression results concerning the relationship between fiscal space and the fiscal policy response to distress. The first, Italy in the global financial crisis (the top panel), is one where a high-debt country swung strongly to fiscal contraction after a crisis. The second, Norway in the early 1990s (the bottom panel), is a clear example of the opposite pattern: in this case, a country with low debt ran highly expansionary policy after a crisis.

The two cases shown in figure 6 are ones where debt and fiscal policy both behave relatively consistently throughout the episode. Perhaps more telling are some of the cases where debt and fiscal policy evolved over the course of the episode. Two such cases are shown in figure 7. Ireland (the top panel) began its 2009 crisis with a low debt-to-GDP ratio, and it initially responded to high distress by undertaking extreme spending measures to stabilize its financial system. However, as its debt ratio rose and distress continued, it swung strongly to fiscal contraction. The other

19. For completeness, we have examined the effects of using the prior surplus either in place of or in addition to the prior debt ratio in all the empirical exercises reported in the paper. Throughout, the results are qualitatively similar to what we find here. The prior surplus enters in ways that are statistically and quantitatively significant; when both variables are included, the statistical significance of the debt ratio is reduced somewhat, but it remains marginally to very significant, and it has a quantitatively more important role than the prior surplus; and the null hypothesis that neither variable enters is overwhelmingly rejected.

20. To form estimates of the change in the high-employment surplus, we need an estimate of trend growth by country; that is, we need an estimate of the $\tau \cdot D\bar{y}$ term that we are able to omit in estimating equation 3. We use each country's average growth over the full sample period, 1980:H1–2017:H2, as our estimate of the growth rate of potential output in the country. For the countries for which we do not have GDP data for the full period, we use the average growth rate over the period for which we have data.

Figure 6. Financial Distress, Debt, and the High-Employment Surplus in Episodes of High Distress: Two Conforming Cases with Consistent Debt[a]

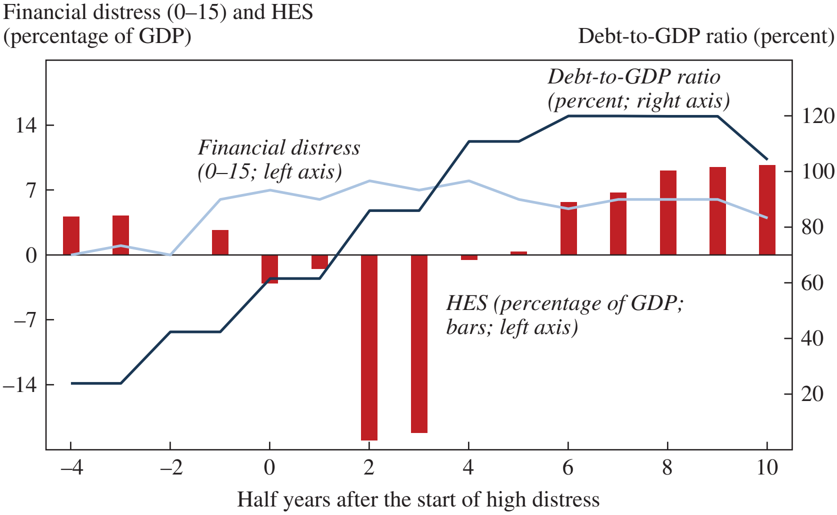


Source: Authors' calculations.

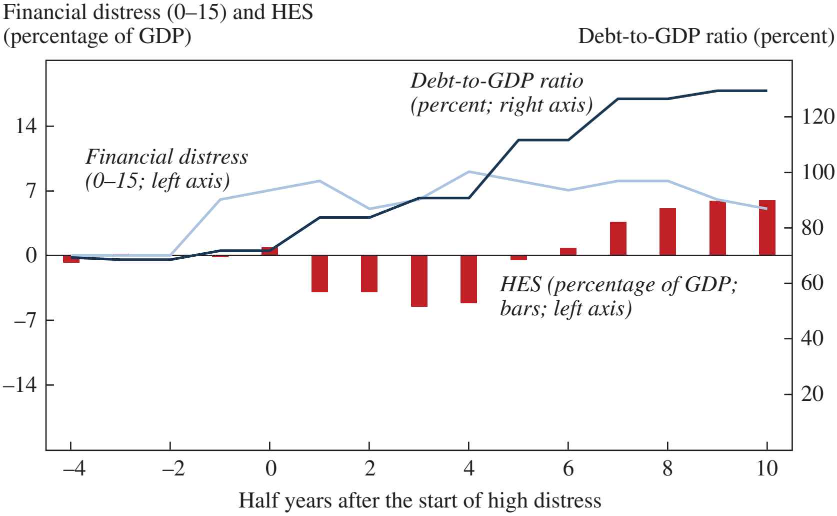
a. HES = high-employment surplus; H = half year. This figure shows the behavior of financial distress, the change in the estimated high-employment surplus from its value two half years before distress reached 7 or more, and the debt-to-GDP ratio in two episodes of high financial distress. The top panel shows a case where debt was high throughout the episode; the bottom panel shows a case where it was low throughout the episode.

Figure 7. Financial Distress, Debt, and the High-Employment Surplus in Episodes of High Distress: Two Conforming Cases with Evolving Debt[a]

Ireland, 2009:H1



Portugal, 2008:H2



Source: Authors' calculations.

a. HES = high-employment surplus; H = half year. This figure shows the behavior of financial distress, the change in the estimated high-employment surplus from its value two half years before distress reached 7 or more, and the debt-to-GDP ratio in two episodes of high financial distress. The two cases shown are ones where the debt-to-GDP ratio rose over the course of the episode and fiscal policy was first expansionary and then moved in a contractionary direction.

case, Portugal in the 2008 crisis (the bottom panel), shows a similar though less extreme pattern. Because the regressions always consider the recent (but prior) level of the debt ratio, cases where distress continued and fiscal policy swung toward contraction as debt rose fit the regression finding that the policy response to distress is more contractionary when debt is higher.

The cases shown in figures 6 and 7 help ground the regression results. They show that the estimates are consistent with the behavior of debt and the high-employment surplus in several key crisis episodes. But, obviously, not every episode cleanly matches the regression findings. For example, South Korea after its 1997 crisis had ample fiscal space as measured by its debt-to-GDP ratio, but nevertheless pursued austerity. And the United States after its crisis in 2007 is an example of a country with somewhat high debt that nevertheless pursued aggressive fiscal stimulus and financial rescue. Such nonconforming cases are reflected in the standard errors of the regression estimates.

III. Statistical Evidence on Why Fiscal Space Matters for the Policy Response

The previous section shows that the fiscal policy response to financial distress varies dramatically with a country's prior debt-to-GDP ratio. Countries with low debt-to-GDP ratios on the eve of financial distress expand aggressively, while countries with high debt-to-GDP ratios tighten sharply. The obvious question is why.

III.A. Possible Explanations

One possibility is that the link between fiscal policy after crises and the debt-to-GDP ratio reflects variation in sovereign market access. Perhaps investors in government bonds are sensitive to a country's fiscal space. In this case, countries with a higher debt-to-GDP ratio may experience larger rises in interest rates after financial distress, and thus be less able or willing to engage in fiscal expansion. Indeed, in extreme cases, market access may be so constrained that higher-debt countries may find themselves forced to undertake extreme austerity because they are unable to borrow. Lower-debt countries, in contrast, may have better market access and thus be able to run very expansionary policy.

The broad alternative explanation for the link between the fiscal response to a crisis and fiscal space involves policymakers' choices. Perhaps policymakers have views about the desirability of fiscal expansion or austerity that vary with the debt-to-GDP ratio. For example, policymakers may believe that financial rescue and countercyclical stimulus are appropriate when the debt ratio is low, but not when it is high. Likewise, they may believe that postcrisis austerity is called for when the debt load is heavy, but not when it is light. As a result, a higher-debt country might choose to expand little or pursue austerity after a financial crisis, while a lower-debt country may choose to expand aggressively, even if neither country faces pressure from markets.

The policymakers making such choices are not necessarily those within the country. For example, countries in the European Union (or those wishing to join the EU or the euro zone) agree to certain standing rules about debt and deficit levels. Thus, the fiscal response of such countries to a financial crisis may vary with their debt-to-GDP ratio because of the ideas and rules of the EU. Countries with a higher debt-to-GDP ratio may find themselves pressured by the EU to conduct austerity after a crisis, while those with a lower debt ratio remain free to conduct financial rescue efforts and countercyclical stimulus.

The fiscal conditionality imposed by the IMF and other international organizations as part of a bailout reflects a sort of hybrid between the market access and policymaker choice explanations. Countries typically only turn to the IMF when there is an extreme lack of market access. Thus, being subject to IMF conditionality is in a fundamental sense an indicator of severe market constraints. If market access depends on fiscal space, then being subject to IMF conditionality could be thought of as the mechanism by which fiscal space affects the fiscal response to financial distress. However, the nature of the conditionality accompanying the bailout, such as the severity of the required austerity and the speed with which the IMF seeks to return countries to private borrowing, depends on the ideas of IMF policymakers about appropriate fiscal policy, and those may also be affected by countries' fiscal space. In the extreme, a country that is forced to go to the IMF for reasons unrelated to its debt load could nevertheless have its fiscal response be related to its debt-to-GDP ratio purely because of IMF ideas.

The main goal of this section and the next one is to obtain evidence about the relative roles of sovereign market access and policymakers' choices in accounting for the link between the fiscal response to financial distress and the debt-to-GDP ratio. This section considers statistical evidence. As we describe below, there are various direct measures of market access, such as sovereign bond rates and credit ratings. If market access is key, these variables should be better predictors of the fiscal policy response to a crisis than the debt-to-GDP ratio. Indeed, if market access is crucial and the direct measures are good indicators, the debt ratio would have little or no predictive power once these measures are included. In contrast, if market access is not crucial and policymakers emphasize the debt ratio in making choices about fiscal policy, then including direct measures of market access should have little impact on the predictive power of the debt ratio for the policy response.

III.B. Measures of Sovereign Market Access

We consider four relatively direct measures of sovereign market access. The first is the spread on credit default swaps for government debt. Concretely, we use the spread on five-year senior government debt.[21] Because a CDS contract is insurance against default, in principle the CDS spread should be an excellent measure of the premium a country must pay to borrow because of fears about its solvency. Thus, it should be a good measure of market access.

In practice, however, CDS spreads have several drawbacks. First, these contracts did not exist at all until the 1990s, and our data for most countries do not begin until 2004. Second, the markets are often thin, and in some cases inoperative. For example, the reported CDS spread for Greece shows literally no change from February 2012 to March 2017. Third, the observation for Greece over this period (14,904.36 basis points) is so extreme that using the raw data would effectively amount to just including a dummy variable for Greece in this period. We therefore drop Greece from regressions that include the CDS spread. Finally, the CDS spread on a bond of one specific maturity is an imperfect measure of a government's access to bond markets at other maturities, and CDS spreads omit some important risks that lenders face, notably restructuring designed to not trigger CDS contracts and the inflating away of debt.[22]

21. The data are originally from Credit Market Analysis and Thomson Reuters, downloaded from DATASTREAM on December 3, 2018. We use the average of the daily observations for the last month of the half year to construct our semiannual observations. We link the Credit Market Analysis and Thomson Reuters data by splicing in the last half year where there is overlap. The contracts are denominated either in euros or dollars, with the exception of those for Japan, which are in yen.

22. We make two adjustments to the CDS data. First, in the handful of cases where the splicing implies a small negative spread, we set the spread to zero. Second, although data for most of the countries in our sample begin in 2004, for some countries whose debt was regarded as extremely safe, they do not begin until later. In order to mitigate somewhat the loss of observations from using the CDS data, we set the spread for these countries from 2004 until it is first available to zero. The result is a series that covers 29 of the 30 countries in our sample for 2004–17. (The missing country is Luxembourg, which did not issue long-term debt during this period.) Using the data without these adjustments yields very similar results.

Our second measure is simpler: long-term interest rates on sovereign debt. In particular, we use the nominal rate on long-term (roughly 10-year) government bonds.[23] These data have the advantages of being available for a large fraction of our sample, of coming from relatively thick markets, and, as with CDS spreads, of reflecting market-based assessments of the riskiness of countries' debt. However, they are affected by factors other than riskiness, notably short-run monetary policy and trend inflation. And, as with the CDS spread, the long-term interest rate on a 10-year bond is an imperfect measure of the premium a government must pay at other maturities.[24]

Our third measure is Standard & Poor's (S&P) sovereign bond ratings. We convert S&P's letter grades into numerical scores by making the step between each grade of equal size, with higher letter grades corresponding to higher scores.[25] This is arguably our most preferred measure of market access: it is continuous, available for most of our sample, and reflects professional assessments of a wide range of information about the riskiness of countries' debt. At the same time, S&P's assessments are necessarily imperfect, market access may not be linear in S&P's letter grades, and no

23. We begin with IMF data on long-term government bond rates (downloaded from International Financial Statistics, data.imf.org/IFS). When these are not available, we use the long-term rates from the OECD (downloaded from Federal Reserve Economic Data, https://fred.stlouisfed.org/), splicing the two series (in levels) at the point of overlap. The data are generally for bonds with maturities of about 10 years. For Mexico until 2001 and Turkey, however, the data are for bonds with maturities of roughly 2 years. We use the observations for the last month of the half year. For the handful of cases where the data for the last month of the half year are not available, we use the observation for the previous month. Finally, although the IMF reports data for Luxembourg, the documentation notes that Luxembourg stopped issuing long-term debt in January 1985, and that the reported data are for private debt (IMF 2018). We therefore treat the observations for Luxembourg starting in 1985 as missing.

24. We also consider the spread between a country's long-term rate and the German rate (or the lower of the German and U.S. rates), rather than the long-term rate. The relationship between the spread and the policy response is slightly stronger than that between the long-term rate and the policy response when we do not include the debt-to-GDP ratio, but slightly weaker when we include the debt ratio. Using the spread rather than the long-term rate has no discernible effect on the relationship between the debt ratio and the policy response.

25. The data are from https://www.capitaliq.com/CIQDotNet/CreditResearch/RenderArticle.aspx?articleId=2094846&SctArtId=460711&from=CM&nsl_code=LIME&sourceObjectId=10686180&sourceRevId=1&fee_ind=N&exp_date=20281001-19:20:54. We use the rating as of the end of the half year. We assign a value of 30 to a AAA rating, 27 to AA, and so on down to 3 to D (although the lowest rating for the countries in our sample is CC, to which our scale assigns a 9). We add a point when the rating is accompanied by a plus and subtract a point when it is accompanied by a minus.

single measure can capture market access over the full range of maturities of a country's bonds.

Our final measure of market access is a dummy variable for whether a country is subject to an IMF standby arrangement or extended fund facility.[26] As discussed above, countries generally turn to the IMF only when they face severe difficulties borrowing in private markets. Thus, being under an IMF program is a strong indicator of very limited market access. At the same time, however, this variable is also an indicator of being subject to IMF policymakers' views. As a result, the interpretation of any estimated impact is inherently complicated.

With all four indicators of market access, we specify the variable so that a larger value corresponds to more access. Specifically, we multiply the CDS spread, the sovereign bond rate, and the IMF dummy (but not the credit rating variable) by –1.

III.C. Does Sovereign Market Access Affect the Fiscal Response to a Crisis?

We begin by considering the predictive power of the various direct measures of sovereign market access for the response of the high-employment surplus to financial distress. These regressions can show if better market access appears to be associated with a more aggressive fiscal response to a crisis.

For this exercise, we estimate regressions analogous to equation 3, with the debt ratio replaced by one of the measures of market access:

$$
\begin{aligned}
(4)\quad (B_{j,t+h} - B_{j,t-1}) - \tau \cdot (y_{j,t+h} - y_{j,t-1}) &= \alpha_j^h + \gamma_t^h + \eta^h M_{j,t} + \beta^h F_{j,t} \\
&\quad + \lambda^h (F_{j,t} \cdot M_{j,t}) + \sum_{k=1}^{4} \xi_k^h M_{j,t-k} \\
&\quad + \sum_{k=1}^{4} \varphi_k^h F_{j,t-k} + \sum_{k=1}^{4} \mu_k^h (F_{j,t-k} \cdot M_{j,t-k}) \\
&\quad + \sum_{k=1}^{4} \theta_k^h (\Delta B_{j,t-k} - \tau \cdot \Delta y_{j,t-k}) + e_{j,t}^h .
\end{aligned}
$$

Here $M_{j,t}$ is a measure of sovereign market access, and the other variables are as before. Our main interest is in the sequence of estimates of l^h, which show how the fiscal response to distress varies with market access.

26. Information on IMF programs is from https://www.imf.org/external/np/fin/tad/extarr1.aspx. Our measure is a dummy variable for whether a country was subject to a program at the end of the half year.

Throughout, the measure of market access in period t is based on information as of the end of the previous half year. Thus, for example, we use information on the long-term interest rate or the S&P rating at the end of the half year before period t in constructing $M_{j,t}$. We are therefore asking whether the fiscal policy response to financial distress varies with the degree of market access the country faced before distress. This timing puts the measures of market access on roughly the same footing as the debt-to-GDP ratio in the regressions presented in the previous section.[27]

One concern is that previous market access may be a less good predictor of current market access than the previous debt ratio is of the current ratio. This could make the estimated interaction effect for the market access variables less comparable to that for the debt ratio in the previous section. As a simple test for this, we regress each market access variable on country and time fixed effects, and four half-yearly lags of itself. The sum of the coefficients on the lag terms range from 0.76 for the IMF dummy to 0.94 for the S&P rating. The comparable sum of the lag terms for the debt ratio is 0.96.[28] The sum of the lag coefficients is overwhelmingly significant in all cases. These findings suggest that there is some validity to this concern, but that it may not cause large incomparabilities. Moreover, we are hesitant to adjust the timing to use the more nearly contemporaneous measures of market access because of endogeneity concerns. Without the lag in timing, it is impossible to distinguish between market access in times of distress affecting the fiscal response to a crisis, and the fiscal response to a crisis affecting the contemporaneous state of market access.

INCLUDING THE MEASURES OF MARKET ACCESS ONE AT A TIME Figure 8 shows the results of estimating equation 4 including the measures of market access one at a time. There are four panels; each corresponds to a different measure of market access. Paralleling the top panel of figure 5, each panel of figure 8 shows the sequence of estimates of the coefficient on the interaction term, scaled for ease of interpretation. As discussed above, we specify the

27. Because the market access measures are high frequency, while the debt ratio is annual, the market access variables are slightly less lagged than the debt variable. For the second observation of each year, the debt variable corresponds to the end of the previous calendar year, whereas the market access variables correspond to the end of June.

28. Because there are not true semiannual observations for the debt data (we simply repeat the value for both half years), we run the regression for the debt-to-GDP ratio at an annual frequency.

measures of market access in such a way that a negative coefficient on the interaction term means that better market access is associated with a smaller value of the change in the high-employment surplus, and hence with a more expansionary fiscal policy response. We also always multiply the coefficients on the interaction term by 7 so that we are focusing on substantial financial distress. For the IMF dummy, we make no further adjustment, so what is shown is simply the difference in the policy response between a country not subject to an IMF program and a country subject to one. For the other three measures (the CDS spread, the long-term interest rate, and the S&P rating), we also multiply the estimated coefficients on the interaction term by twice the standard deviation of the measure, which is analogous to our treatment of the debt-to-GDP ratio in figure 5. In short, each panel shows the estimated *difference* in the response of the high-employment surplus as a percentage of GDP to an innovation in financial distress of 7 in a country with better market access versus a country with worse market access.

The top panel of figure 8 shows that there is basically no correlation between a country's CDS spread and its fiscal response to financial distress. The estimated impact of having a spread that is 2 standard deviations lower (roughly 230 basis points) on the response of policy to an innovation of 7 in financial distress is of irregular sign, quantitatively small, and never close to statistically significant. The confidence intervals include moderately negative and moderately positive values but do not include large values.

For the other three measures, better market access is associated with a more expansionary fiscal response to financial distress. The results are strongest for the IMF program dummy (the bottom panel of figure 8). The point estimates imply that the response of the high-employment surplus in a country that is not subject to an IMF program to an innovation in distress of 7 is more expansionary by up to 6 percent of GDP. The null hypothesis that the responses do not differ between a country not subject to a program and a country subject to one is decisively rejected, with a maximum t statistic over 4.

The results for the other two measures are not as overwhelming, but are still strong. For the interest rate on long-term government debt (the second panel of figure 8), having a sovereign yield 2 standard deviations lower (8.6 percentage points) is associated with a substantially more expansionary fiscal policy response to a financial crisis (often exceeding 2 percent of GDP), but with wide confidence intervals and a maximum

Figure 8. The Relationship between the High-Employment Surplus after a Financial Crisis and Individual Direct Measures of Sovereign Market Access[a]

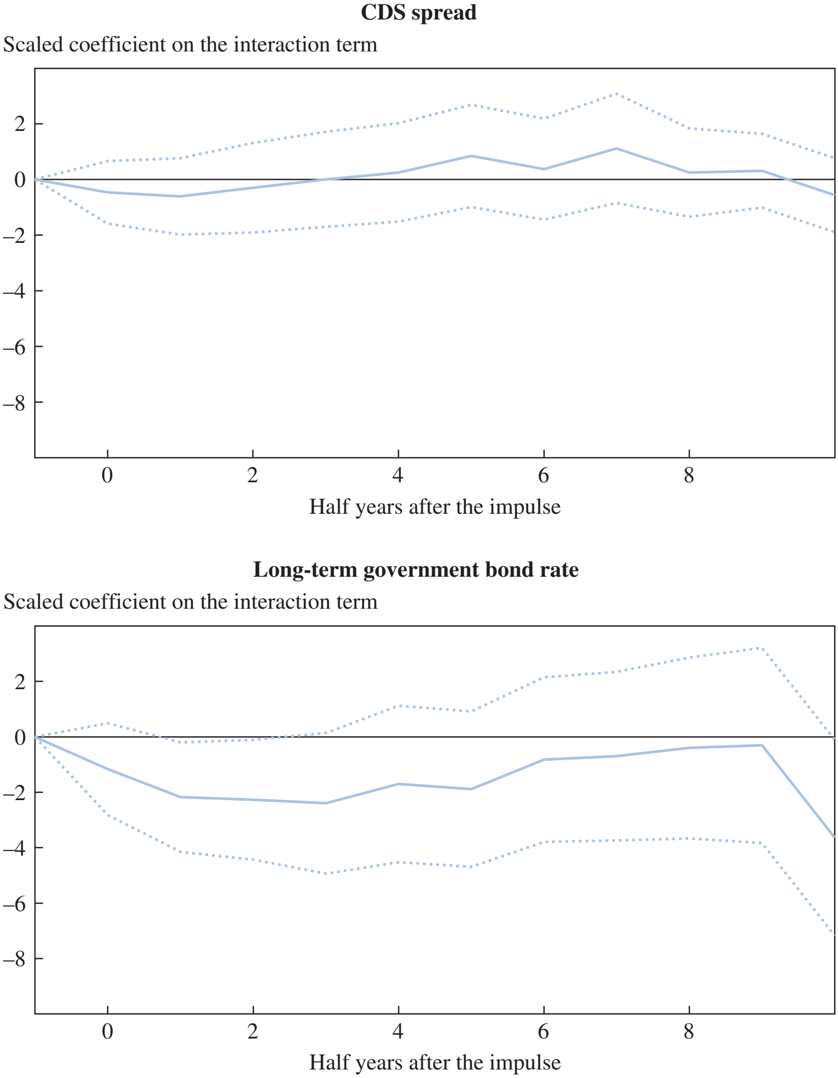
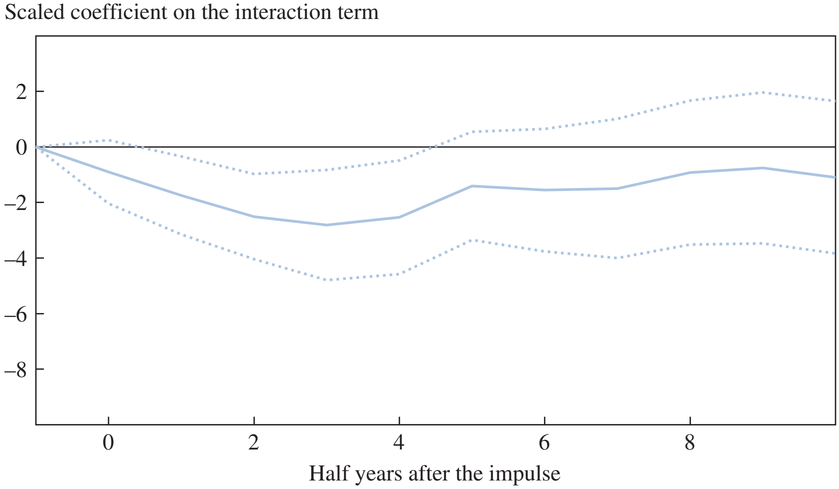
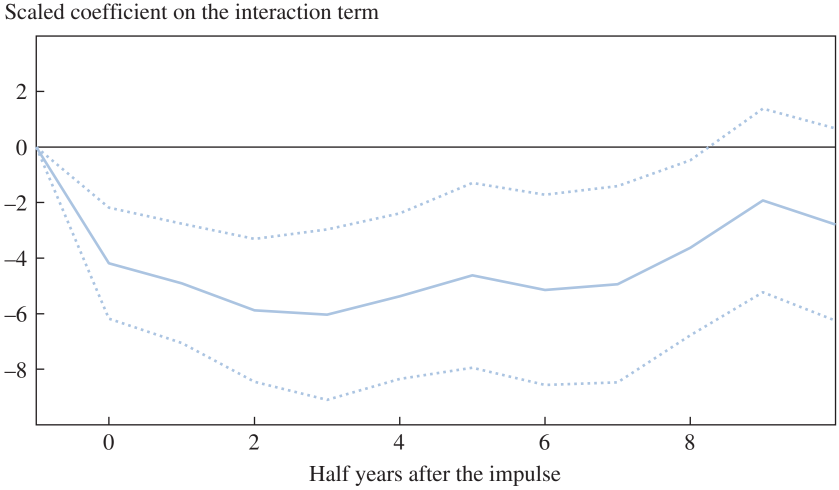


Figure 8. The Relationship between the High-Employment Surplus after a Financial Crisis and Individual Direct Measures of Sovereign Market Access[a] (*Continued*)

Standard & Poor's rating



IMF program



Source: Authors' calculations.

a. The panels show how the response of fiscal policy to financial distress varies with individual direct measures of sovereign market access. Each panel is based on estimates of equation 4 over the period 1980:H1–2017:H2 using a different measure of market access; H = half year. The top three panels show the estimated values of λ^h for different values of h, scaled by 7 times twice the sample standard deviation of the relevant measure of market access. Thus, they show how the response to an innovation of 7 in financial distress changes with an improvement of 2 standard deviations in the indicated measure of access. The bottom panel shows the estimated values of λ^h for different values of h, scaled by 7. Thus, it shows how the response differs between a country not subject to an IMF program and a country subject to a program. The dotted lines show the 2-standard-error confidence bands.

t statistic of only 2.2.[29] And the third panel shows that a 2-standard-deviation improvement in a country's S&P rating (roughly 8 points on our numerical scale—for example, from B to A−) is associated with a more expansionary policy response of similar size (2 percent of GDP or more) and is highly statistically significant at short horizons, with a maximum *t* statistic of 3.3.

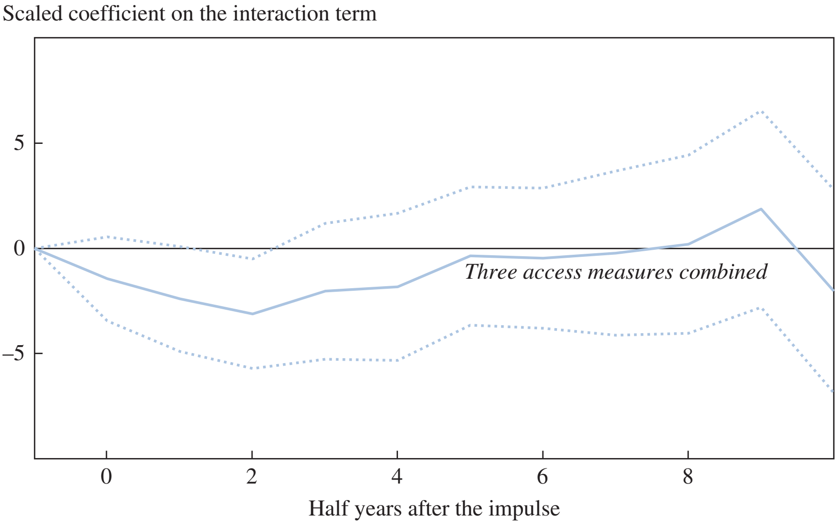
INCLUDING MULTIPLE MEASURES OF MARKET ACCESS Each of the direct measures of market access captures slightly different features of a country's ability to borrow. Moreover, each measure has unique strengths and limitations. Thus, it is sensible to see if the measures considered jointly have more substantial predictive power for the fiscal response to a financial crisis than each considered separately. To do this, we expand equation 4 to include the level and interaction with distress (as well as the appropriate lags) of three measures of market access: the long-term government bond rate, the S&P sovereign rating, and the dummy for being under an IMF program. We exclude the CDS spread because its inclusion limits the sample so severely.

For each horizon of the impulse response function, we consider the point estimate and statistical significance of the sum of the coefficients on the interactions of each measure of market access with financial distress times twice its standard deviation. As in figure 8, we also multiply the weighted sum of the estimated coefficients on the interaction terms by 7. Thus, it shows how the response to an innovation of 7 in financial distress changes with an improvement of 2 standard deviations in all three measures of market access.

Figure 9 shows the results. The point estimates of the weighted sum of the interaction coefficients are negative for the first two and a half years after a crisis. This suggests that countries with better market access engage in more expansionary fiscal policy after a crisis than countries with worse market access. The point estimates after two and a half years are very close to zero (until year 5, when they fluctuate substantially). Perhaps more important than the point estimates is the statistical significance of the weighted sum. The standard errors of the sum of the interaction effects are substantial, particularly at longer horizons. As a result, the impact of market access on the fiscal response to a crisis is statistically significant at only one horizon—horizon 2. That the three measures of

29. The value we use for the standard deviation of the long-term interest rate excludes Turkey. Our data for the long-term rate include only a handful of values for Turkey, but not enough consecutive data points for any observations from Turkey to enter the regression. In addition, the values of the long-term rate for Turkey are so extreme (generally close to 100 percent) that including them would almost double the overall standard deviation.

Figure 9. The Relationship between the High-Employment Surplus after a Financial Crisis and Multiple Direct Measures of Sovereign Market Access[a]



Source: Authors' calculations.

a. This figure shows how the response of fiscal policy to financial distress varies with sovereign market access when several measures of market access are considered jointly. It is based on estimates of equation 4 over the period 1980:H1–2017:H2 (H = half year) including three measures of market access: the long-term government bond rate, the Standard & Poor's rating, and a dummy variable for not being under an IMF program. The figure plots the sum of the interaction coefficients for each measure of market access times twice its standard deviation. The sum is then multiplied by 7, so that it shows how the response to an innovation of 7 in financial distress changes with an improvement of 2 standard deviations in all three measures of market access. The dotted lines show the 2-standard-error confidence bands.

market access considered jointly have a less precisely estimated impact on the fiscal response to a crisis than two of the measures considered individually (the S&P rating and the IMF dummy) reflects both changes in the sample caused by including all three measures and interactions between the various measures.

III.D. Does the Debt Ratio Affect the Fiscal Response through Market Access?

The previous regressions show whether direct measures of sovereign market access predict the fiscal policy response to a financial crisis. They do not, however, answer the question raised by the results in section II, which is why the debt-to-GDP ratio appears to matter. This issue can be addressed by testing whether including direct measures of market access attenuates the predictive power of the debt ratio.

INCLUDING THE MEASURES OF MARKET ACCESS ONE AT A TIME To do this, we estimate regressions along the lines of equations 3 and 4, but now including both (the negative of) the debt ratio and a measure of market access:

$$
\begin{aligned}
(5)\quad (B_{j,t+h} - B_{j,t-1}) - \tau \cdot (y_{j,t+h} - y_{j,t-1}) = {} & \alpha_j^h + \gamma_t^h + \vartheta^h S_{j,t} + \eta^h M_{j,t} + \beta^h F_{j,t} \\
& + \delta^h (F_{j,t} \cdot S_{j,t}) + \lambda^h (F_{j,t} \cdot M_{j,t}) \\
& + \sum_{k=1}^{4} \rho_k^h S_{j,t-k} + \sum_{k=1}^{4} \xi_k^h M_{j,t-k} \\
& + \sum_{k=1}^{4} \varphi_k^h F_{j,t-k} + \sum_{k=1}^{4} \omega_k^h (F_{j,t-k} \cdot S_{j,t-k}) \\
& + \sum_{k=1}^{4} \mu_k^h (F_{j,t-k} \cdot M_{j,t-k}) \\
& + \sum_{k=1}^{4} \theta_k^h (\Delta B_{j,t-k} - \tau \cdot \Delta y_{j,t-k}) + e_{j,t}^h .
\end{aligned}
$$

Thus, this regression includes both the interaction between financial distress and the debt ratio and the interaction between distress and a direct measure of market access. We first include the measures of market access one at a time.

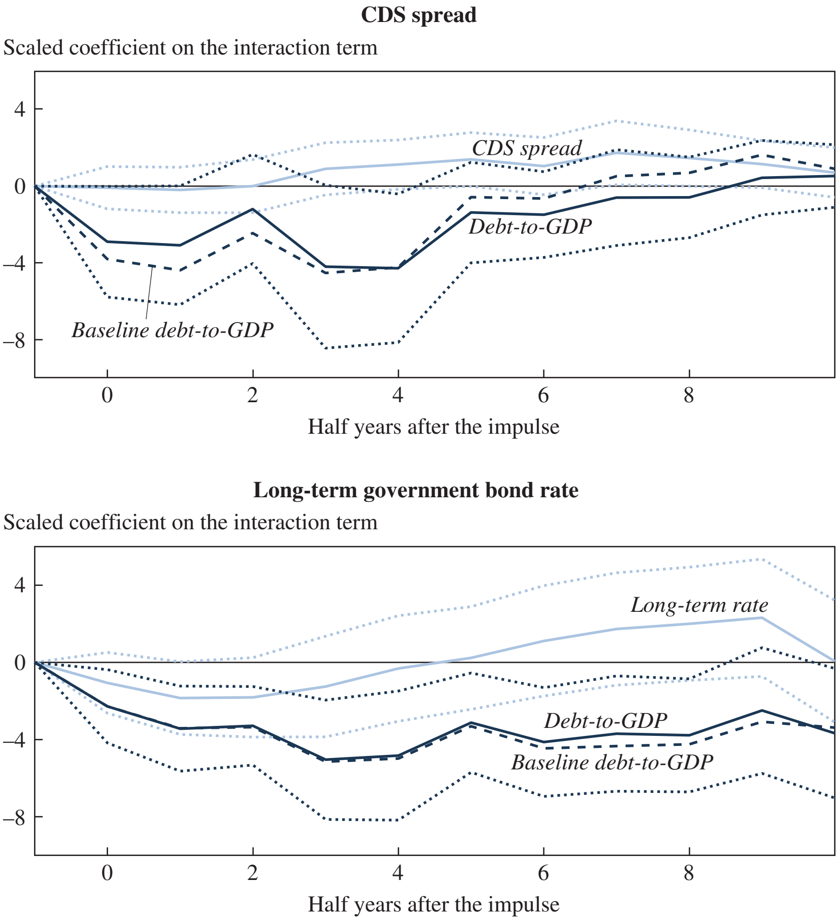
Figure 10 shows the results. Each panel considers a different measure of market access. The panels show two sets of estimates of differences between the fiscal policy responses to an innovation of 7 in financial distress: between countries with better and worse market access (with the differences measured in the same way as in figure 8), and between countries with smaller and larger debt-to-GDP ratios (with the difference equal to 2 standard deviations, or roughly 70 percentage points, as in the top panel of figure 5). Because the key question is whether including the measures of market access weakens the predictive power of the debt ratio, we also include the scaled coefficient on the interaction term for the debt ratio from the specification that excludes the market access measures (equation 3). The sample used to estimate the specification excluding the market access measure is adjusted to match that for the specification including the market access measure in each case. Thus, the two estimates of the interaction effect for the debt ratio given in a panel differ only because of the inclusion of the market access measure. The differences in the sample also explain why the estimate of the debt interaction effect excluding market access differs across the panels and from the baseline results given in the top panel of figure 5. To keep the figures readable, we do not show the interaction effect for the market access variable from the specification excluding the debt ratio (equation 4), but we do discuss how this effect changes as well.

The basic message of figure 10 is that including the direct measures of sovereign market access attenuates the estimated effect of the debt-to-GDP ratio on the fiscal response to a crisis by at most a moderate amount (and usually just a very small amount), while including the debt ratio often has a large impact on the estimated effects of the measures of market access. The top panel of figure 10 considers the case where the CDS spread is used to measure market access. Including the interaction with the CDS spread reduces the interaction with the debt ratio by a moderate amount at very short horizons, but increases it at longer horizons. That is, fiscal policy after a financial crisis is somewhat less responsive to the debt ratio when the CDS spread is included, but only at short horizons. And even in this smaller sample, the impact of the debt ratio on the fiscal response at short and medium horizons is large, though only marginally significant. In contrast, including the debt ratio makes the results for the CDS spread even weaker than before. The estimates are generally wrong-signed, and at long horizons (with, as just noted, small sample sizes), marginally significant.

The second panel of figure 10 shows that including the interest rate on long-term government debt has essentially no effect on the estimated impact of the debt-to-GDP ratio on the fiscal policy response to financial distress. It also shows that including the debt ratio noticeably weakens the predictive power of the long-term rate for the policy response. The estimated impact remains negative at short horizons, but it is now at most weakly significant; and it is now positive, though statistically insignificant, at longer horizons. The third panel of the figure shows that when the S&P rating is included, the estimated impact of the debt ratio on the fiscal response to a crisis weakens by about 20 percent, but it remains large and often statistically significant. It also shows that including the debt ratio has a larger effect on the estimated impact of the S&P rating. Its estimated effect remains consistently negative, but it is somewhat smaller than before and is statistically significant only at horizon 2.

Finally, the bottom panel of figure 10 shows the results for the dummy for being subject to an IMF program. The inclusion of a measure of market access again reduces the estimates of the effect of the debt-to-GDP ratio

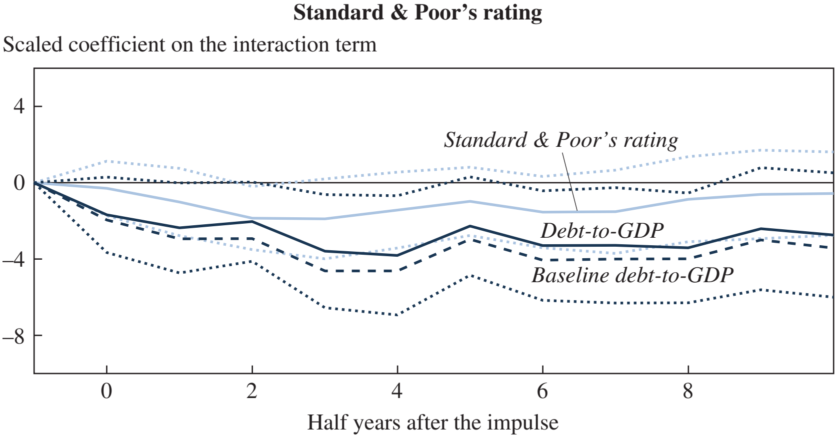
Figure 10. The Relationship between the High-Employment Surplus after a Financial Crisis and Both Individual Direct Measures of Sovereign Market Access and Fiscal Space[a]

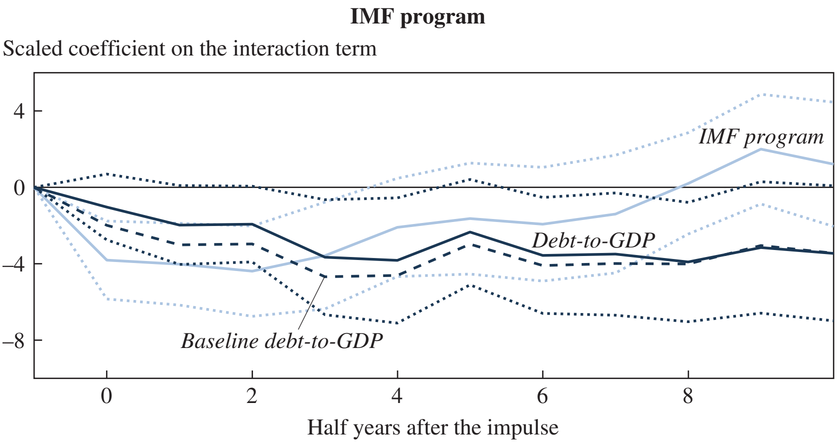


(*continued on next page*)

on the response to financial distress by about 20 percent, though they are still large and are often significant. And again, the inclusion of the debt ratio has a larger effect on the estimates of the effects of the measure of market access. At short horizons, the estimated effect continues to be large and overwhelmingly significant, though it is noticeably weaker than before. But at medium horizons, the effects move from very large and highly significant to moderate and not significant; and at long horizons, they become positive, though not significant.

Figure 10. The Relationship between the High-Employment Surplus after a Financial Crisis and Both Individual Direct Measures of Sovereign Market Access and Fiscal Space[a] (*Continued*)





Source: Authors' calculations.

a. The panels show how the response of fiscal policy to financial distress varies with individual direct measures of sovereign market access, based on a specification that also includes the debt-to-GDP ratio and its interaction with financial distress. Each panel is based on estimates of equation 5 over the period 1980:H1–2017:H2 using a different measure of market access; H = half year. The top three panels show the estimated values of λ^h for different values of h, scaled by 7 times twice the sample standard deviation of the relevant measure of market access. Thus, they show how the response to an innovation of 7 in financial distress changes with an improvement of 2 standard deviations in the indicated measure of access. The bottom panel shows the estimated values of λ^h for different values of h, scaled by 7. Thus, it shows how the response differs between a country not subject to an IMF program and a country subject to a program. Each panel also shows the estimated values of δ^h for different values of h, scaled by 7 times twice the sample standard deviation of the debt ratio. Thus, it shows how the response changes with a decline of 2 standard deviations in the debt-to-GDP ratio in the specification including the indicated measure of market access and its interaction with financial distress. The line denoted "baseline debt-to-GDP" shows how the response changes with a decline of 2 standard deviations in the debt-to-GDP ratio in the specification excluding the measure of market access (but using the same sample as the expanded regression). The dotted lines show the 2-standard-error confidence bands.

INCLUDING MULTIPLE MEASURES OF MARKET ACCESS We again consider a more extreme test. We allow the fiscal policy response to financial distress to depend on the debt-to-GDP ratio and on three measures of sovereign market access: the long-term government bond rate, the S&P rating, and the IMF dummy. That is, we expand equation 5 to include three measures of market access (and their interactions with financial distress), rather than just one. (We again do not include the CDS spread, on the grounds that doing so would entail a very large reduction in the sample size and that the previous results find essentially no effect of the spread.)

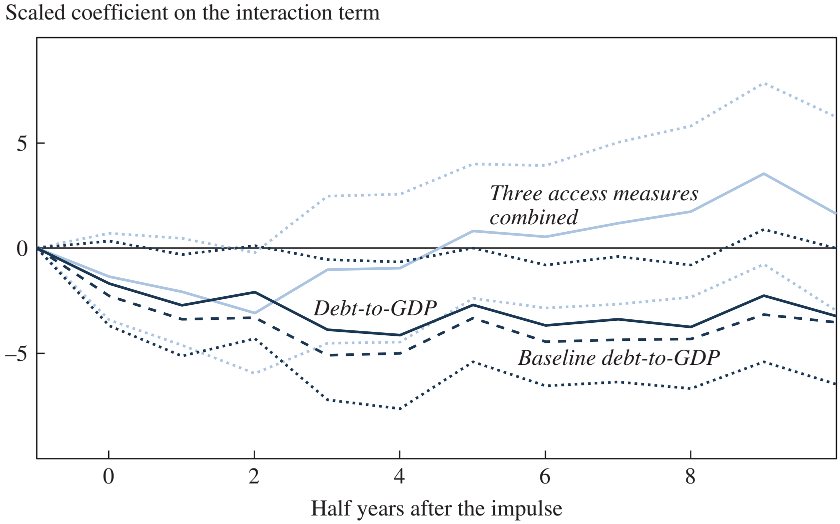
Figure 11 shows two scaled estimates of the interaction effect with financial distress for various horizons: that for the debt ratio, and that for the weighted sum of the three direct measures of market access (calculated as in figure 9). As in figure 10, we also show the interaction effect for the debt ratio from equation 3 (where the measures of market access are excluded), but for the same sample as those from expanded specification.

Figure 11 shows that the estimated sensitivity of the fiscal response to a crisis to the debt-to-GDP ratio is only mildly affected by the simultaneous inclusion of multiple direct measures of market access. The impact remains large and generally statistically significant; the inclusion of the multiple measures of market access reduces the point estimates by about 20 percent relative to our baseline estimates. Thus, the debt-to-GDP ratio appears to matter substantially for the fiscal response to a crisis, even when one takes into account the behavior of the three direct measures of market access. As in figure 9, the weighted sum of the interaction coefficients for the three measures of market access is initially negative, but it is only marginally significant at one horizon—horizon 2. The weighted sum of the interaction effects then turns positive, with very large standard errors.

III.E. Discussion

The statistical results advance our understanding of the fiscal policy response to financial distress in two ways. First, they provide some evidence that market access matters. There is a moderately large, moderately significant, and relatively robust relationship between the rating on a country's sovereign debt and its fiscal policy response to distress; and there is a large, highly significant, and very robust relationship between a country being subject to an IMF program and its short-run fiscal response. Countries with higher-rated debt respond more aggressively to distress, and the immediate response of countries that are subject to an IMF program is far more contractionary than that of countries that are not. In addition, countries respond less aggressively to distress when the interest rate on

Figure 11. **The Relationship between the High-Employment Surplus after a Financial Crisis and Both Multiple Direct Measures of Sovereign Market Access and Fiscal Space**[a]



Source: Authors' calculations.

a. This figure shows how the response of fiscal policy to financial distress varies with measures of sovereign market access when several measures of market access are considered jointly, and the specification also includes the debt-to-GDP ratio and its interaction with financial distress. The figure is based on estimates of equation 5 over the period 1980:H1–2017:H2 (H = half year) including three measures of market access: the long-term government bond rate, the Standard & Poor's rating, and the dummy variable for not being under an IMF program. The figure plots the sum of the interaction coefficients for each measure of market access times twice its standard deviation. The sum is then multiplied by 7 so that it shows how the response to an innovation of 7 in financial distress changes with an improvement of 2 standard deviations in all three measures of market access. The figure also shows how the response changes with a decline of 2 standard deviations in the debt-to-GDP ratio in the specification including the three measures of market access and their interaction with financial distress. The line denoted "baseline debt-to-GDP" shows how the response changes with a decline of 2 standard deviations in the debt-to-GDP ratio in the specification excluding the measures of market access (but using the same sample as the expanded regression). The dotted lines show the 2-standard-error confidence bands.

their long-term debt is higher, and they respond less aggressively at longer horizons when they are subject to an IMF program. However, these relationships are not as consistent as those for the S&P rating and for the immediate impact of being under an IMF program.

Second, we find that including direct measures of sovereign market access does not eliminate the estimated impact of the debt-to-GDP ratio on the fiscal response to a financial crisis. Indeed, using a range of measures (both singly and in combination) of market access attenuates only slightly the correlation between the debt-to-GDP ratio and the fiscal policy response. This suggests that an important part of the relationship between the debt ratio and the fiscal response to a crisis operates through channels other than market access. The natural alternative is that those other channels involve choices by policymakers.

IV. Narrative Evidence on the Motivation for the Fiscal Policy Response to Crises

In this section, we turn from statistical evidence to narrative evidence. There are several reasons that the broader, more qualitative evidence provided by narrative sources may be able to provide important additional information about the determinants of the fiscal policy response to financial distress. First, and most obviously, our interest is in the motivations for policy actions. Thus, there is likely to be valuable evidence from policymakers' statements, news accounts, and other descriptions of the policymaking process available to informed contemporary observers.

Second, all our direct measures of sovereign market access are imperfect. As a result, the statistical relationship between the measures and the fiscal policy response to financial distress may understate the importance of market access. It is therefore useful to examine whether analysts making more holistic judgments about the drivers of policy decisions perceived a greater role for problems with market access than comes through in our statistical analysis.

Third, and closely related, problems with market access may involve large and complicated nonlinearities that would be extremely difficult to detect statistically. For example, suppose that there is some level of fiscal expansion that would trigger a sudden, sharp rise in interest rate spreads and severe problems with market access, and that this level varies across situations. If policymakers are aware of these limits and are careful not to breach them, concerns about a possible imminent loss of market access could be driving fiscal policy without showing up in our measures. But such considerations might be apparent to observers monitoring countries' decisionmaking.

Motivated by these considerations, in this section we seek to provide narrative evidence on the relative importance of market access and policymakers' choices in driving fiscal policy actions during and after crises. We also investigate the degree to which government debt ratios appear to underlie or influence market access and policymakers' choices. The narrative analysis focuses on the 22 episodes of high distress in our sample.

IV.A. Source and Approach

Our source for this analysis is the Economist Intelligence Unit's *Country Reports*.[30] The EIU is a subscription service providing information on political and economic developments in a wide range of countries. The reports are aimed at investors and other market professionals, and they appear to be of reasonably high quality. Each one is about 25 pages, and makes frequent reference to data, legislative debates, and political developments. The service relies, in part, on the reporters and analysts working for the *Economist* magazine for its information. Because the reports focus particularly on policy actions, we find them to be a plentiful source for information on the motivations for those actions.

The EIU *Reports* are available quarterly until early 2000, and monthly thereafter. The reports for the first quarter (which came out in early January) and the third quarter (which came out in early July) are roughly comparable in timing to the *OECD Economic Outlook*, from which we derive our measure of financial distress (the *Economic Outlook* generally comes out in December and June). When the reports become monthly, we treat those for January and July as the parallel to the OECD volumes.[31] Our approach is to read the EIU *Reports* for a country corresponding in timing to the OECD *Economic Outlook* starting in the half year before financial distress reached 7 or above. We read nine issues for each episode of high distress—two a year from one half year before the crisis to four years after.

We look for what the EIU *Reports* say about four questions:

1. What is the current and/or prospective stance of fiscal policy?
2. What is the motivation given for the fiscal developments?
3. Does the EIU mention the debt-to-GDP ratio as a concern or as an underlying motivation?
4. Is there anything else of note relevant to fiscal policy actions?

30. The EIU *Reports* after 1996 are available from EIU.com; those before 1996 are available from the Economist Intelligence Unit's Country Reports Archive through ProQuest. Binder (2018) uses the *Country Reports* as a narrative source to study political pressure on central banks.

31. We do this in part for consistency with the early years of the EIU *Reports*, and in part because, for a transition period, the December and June issues are updates rather than full reports. In addition, there are a few irregularities in which EIU reports we consider. For example, the report for Sweden for 1994:Q1 is not available from ProQuest. Similarly, in a few cases in the early 2000s, the January and July issues are updates and the December and June issues are full reports. In such cases, we use a rule of reason and read the obvious alternative report.

From the answers to question 1, we identify whether fiscal policy at the time was perceived by the EIU to have been on net expansionary or contractionary, and whether the net movement was large or small. The overall direction of fiscal policy is usually stated quite directly. To categorize size, we rely mainly on adjectives. Fiscal moves described with words like "mild," "minor," and "small" are classified as small; those described with words like "enormous," "unprecedented," and "extreme" are classified as large. This scaling is obviously rough. In the tables that summarize the narrative evidence, we therefore note cases that are marginal for a category.

In classifying the motivation given by the EIU for fiscal actions (question 2), we are open to nuance and multiple motivations. For fiscal expansions, we identify three possible motivations:

a. Financial rescue.
b. Countercyclical stimulus.
c. Politics.

The EIU typically says why some fiscal action was taken, not why some other action was not taken. That is, the counterfactual is implicitly or explicitly neutral fiscal policy. As a result, market access is inherently not an explanation for fiscal expansions, and all the possible motivations correspond to policymakers' choices. Actions taken for countercyclical stimulus or financial rescue can be thought of as reflecting policymakers' ideas: policymakers take these actions because they believe the policies will be helpful. "Politics" (category c) refers to fiscal stimulus undertaken to try to help win an election.

For fiscal contractions, we consider five possible motivations:

d. Market access.
e. Conditionality imposed as part of a bailout.
f. Policymakers' ideas.
g. European Union fiscal rules.
h. Countercyclical austerity.

"Market access" captures austerity that the EIU identifies as being driven by very high sovereign spreads, inability to borrow, or other problems with private sovereign funding. The final three motivations correspond to policymakers' choices. As described above, policymakers' ideas and EU rules largely differ in *whose* ideas are mattering—those of domestic policymakers or EU officials. "Countercyclical austerity" (category h) refers to contractionary fiscal actions taken to prevent overheating, and can again be thought of as a special form of policymakers' ideas. Finally, "conditionality" is the hybrid of market access and IMF (or other bailout organization) choices discussed above.[32]

Question 3 reflects both information about policymakers' concerns about the debt-to-GDP ratio in an episode and the EIU's perspective. We take note of the EIU's assessment on the grounds that when we do not have direct information on policymakers' views, it at least provides the assessment of one contemporary informed observer of whether debt was important. However, we put the main emphasis on the information that the EIU provides about policymakers' views. We pay particular attention to whether the EIU believes that the debt-to-GDP ratio is driving the proximate motivation—for example, whether it cites the debt-to-GDP ratio as something that is affecting market access or policymakers' ideas.

Online appendix B provides our detailed notes and selected quotations for each EIU country report that we read, organized around the four questions discussed above. Table 2 provides the topline summary of what the EIU says about the net direction, size, and motivation for fiscal actions in each of the 22 episodes of high financial distress.[33]

IV.B. Postcrisis Fiscal Expansions

Table 3 summarizes the narrative evidence from the EIU on net fiscal expansions after high financial distress. According to the EIU, in almost all 22 crisis episodes in our sample, there was at least a short period of deliberate net fiscal expansion after the start of high distress. The only exceptions are Mexico (in the mid-1990s) and Hungary (in 2009). In both cases, high financial distress followed extreme exchange rate fluctuations, and the country was already participating in an IMF stabilization program at the time of severe problems in the financial sector. South Korea followed a similar pattern, with very early austerity as part of an IMF program to deal with its 1998 crisis. However, unlike Mexico and Hungary, South Korea undertook fiscal expansion after conditions stabilized. The much more typical pattern is for the fiscal expansions to occur early in the postcrisis period.

32. Not surprisingly, the motivations given in the EIU *Reports* occasionally do not fit into one of the eight categories we identify. In these cases, we note an "other" motivation.

33. As noted above, we start our reading of the EIU *Reports* for each episode one half year before financial distress reached 7 or more. For most episodes, we use the precrisis report only as background to our reading of the later reports. The only cases where we use the precrisis report to characterize the fiscal policy response to financial distress is when the EIU was explicit that fiscal policy actions were motivated by distress, even though the distress had not yet reached 7 or above on our scale. The two cases where this occurs are the 1993 crises in Finland and Sweden.

Table 2. Summary of the Motivations for Fiscal Actions after Financial Crises, as Described in the Economist Intelligence Unit's *Country Reports*[a]

(Crisis episodes are in chronological order; the date given is the half year in which financial distress first reached 7 or above)

United States, 1990:H2
Fiscal policy in the United States was moderately contractionary throughout almost all of the episode. It was driven by two ideas of domestic policymakers. One was that deficits and debt were bad and that the deficit needed to be reduced. The other was that short-run stabilization was best left to monetary policy, which freed fiscal policy to focus on the long-run fiscal outlook and other long-run issues. There was a tiny bit of fiscal expansion in 1992 driven by countercyclical and political motivations.

Norway, 1991:H2
Norway undertook moderately expansionary fiscal policy until well after the crisis for countercyclical purposes and financial rescue. Policy then turned mildly contractionary. The shift was partly the result of policymakers' belief that fiscal rectitude was desirable (despite Norway's extremely strong fiscal position), and partly a countercyclical response to strong growth.

Finland, 1993:H1
Finland initially undertook a large fiscal expansion for financial rescue. Otherwise, fiscal policy was contractionary throughout, often strongly so. The first contractionary moves stemmed from policymakers' beliefs that deficit reduction and low debt were beneficial. Later, a desire to join the Economic and Monetary Union was also important. Market access was mentioned, but was confined to two reports and emphasized only in one.

Sweden, 1993:H1
There was much fiscal expansion in Sweden for financial rescue early on. This expansion was partly counterbalanced by austerity on other fronts owing to a mixture of ideas (the Conservative-led government wanted to shrink the welfare state) and concern about spreads (so a form of market access). Fiscal policy then switched to net austerity. Ideas played a role after the change in government in 1994: the new government also supported getting the budget under control and the electorate seemed to support that. Later on, a desire to meet EU criteria was an important motivation for continued austerity.

Mexico, 1996:H1
There was substantial financial rescue in Mexico, but on net fiscal policy was contractionary throughout, at times strongly so. The main motivation for the austerity was a blend of market access and IMF conditionality. Market access was a substantial problem and the government was trying to regain the confidence of investors and satisfy the IMF. There was also an element of ideas at the end of the period. The outgoing president wanted to prevent a crisis at the handover of government in 2000, and so wanted to run careful policy to prevent trouble.

Japan, 1997:H2
Japan engaged in at least modest fiscal expansion throughout the postcrisis period. Initially, it undertook some fiscal stimulus for countercyclical reasons and financial rescue (though ideas about the importance of fiscal rectitude may have limited actions). In mid-1998, a new government came in and did more aggressive stimulus and financial rescue. There was a little concern about market access that may have restrained stimulus, but it did not lead to austerity. Political considerations and countercyclical aims were the main motivations for continued fiscal expansion later in the postcrisis period.

Table 2. Summary of the Motivations for Fiscal Actions after Financial Crises, as Described in the Economist Intelligence Unit's *Country Reports*[a] (*Continued*)

South Korea, 1997:H2
South Korea initially engaged in some fiscal contraction due to IMF conditionality and problems with market access. Pretty quickly there was a move toward modest fiscal expansion driven by countercyclical aims, financial rescue, and political considerations. Eventually the government shifted to roughly neutral fiscal policy as the economy recovered.

Turkey, 2001:H1
Turkey engaged in substantial financial rescue, especially early on. But this was followed by strong austerity. It is hard to separate market access and IMF conditionality as the motivation for the austerity. Lack of market access led to the IMF program; once in place, the IMF conditionality was strictly enforced (both by the IMF and by markets). At times politics and ideas led policymakers to suggest less austerity. Later in the analysis period, Turkey's desire to join the EU was also a driving force for fiscal contraction.

United States, 2007:H2
Fiscal policy in the United States was initially highly expansionary for countercyclical purposes and financial rescue. It then leveled out and finally turned slightly contractionary. Policymakers' ideas about the harms of the debt and deficits in their own right, plus some concerns about market access (specifically, long-term interest rates and bond ratings), limited the expansionary actions and prompted the ultimate move to contraction.

Iceland, 2008:H1
Iceland initially engaged in a large fiscal expansion related to financial rescue. Then, there was substantial fiscal contraction due to loss of market access and subsequent IMF and Nordic conditionality. The IMF, however, did not force much contraction in 2009, when the recession was at its worst. There was a small continuing role for market access in fostering austerity because ratings on sovereign debt affected borrowing costs of municipalities and companies. Also there may have been a small role for domestic policymakers' own ideas—at times the government seemed to go further with austerity than the IMF required. In addition, there was a fight within the governing coalition, suggesting again that ideas mattered.

United Kingdom, 2008:H1
The United Kingdom spent a substantial amount on financial rescue, but undertook only a very small amount of conventional fiscal stimulus. With David Cameron's election in May 2010, the government switched to extreme austerity. The government said that future market access was the reason, but the EIU emphasized that actual market access was excellent. This juxtaposition may suggest an important role for ideas.

Austria, 2008:H2
Fiscal policy in Austria was initially mildly expansionary for countercyclical reasons and financial rescue, but there was then a switch to substantial austerity. The change appeared to result roughly equally from EU rules and from domestic policymakers' ideas about the benefits of fiscal rectitude.

France, 2008:H2
France initially undertook mildly expansionary fiscal policy for countercyclical purposes and financial rescue. But the size was limited, and policy soon switched to austerity and then remained mildly contractionary. The limitations on the size of the initial expansion and the initial move to austerity were driven mainly by domestic policymakers' ideas about the benefits of fiscal restraint. But market access issues and EU rules both played nontrivial roles in the continued austerity.

(*continued on next page*)

Table 2. Summary of the Motivations for Fiscal Actions after Financial Crises, as Described in the Economist Intelligence Unit's *Country Reports*[a] (*Continued*)

Italy, 2008:H2
Italy engaged in only trivial fiscal stimulus and little financial rescue in the postcrisis period. This was followed by moderate austerity for a number of years. A key reason for the austerity was problems with market access and rising sovereign spreads. EU rules and pressure were another factor.

Norway, 2008:H2
Policymakers in Norway undertook moderate fiscal expansion in response to the crisis, motivated by both countercyclical aims and financial rescue. The EIU thought Norway's oil revenues and low debt were reasons it could do this. The government then scaled back the stimulus in response to recovery to prevent the economy from overheating, but it never switched to net austerity.

Portugal, 2008:H2
Portugal initially engaged in mildly expansionary policy for countercyclical purposes (with just a small amount of financial rescue). Political considerations in the run-up to the election also played a role in driving fiscal expansion. EU rules and actions were a factor in the early moves toward austerity. Starting in mid-2010, Portugal lost the confidence of foreign investors. A lack of market access followed by IMF conditionality led to extreme austerity. The austerity and conditionality continued through the end of the analysis period.

Spain, 2008:H2
Spain initially engaged in mildly expansionary fiscal policy for countercyclical reasons and financial rescue. But policy then turned strongly contractionary and remained so. The main reason was market access, as Spain faced high sovereign spreads and ratings downgrades. Toward the end of the period, the problems with market access led to the possibility of Spain needing to turn to outside help, which created additional pressures for austerity. There were some indications of a role for domestic policymakers' ideas, but those were never central.

Sweden, 2008:H2
Despite its comparatively small-government, supply-side ideology, the Swedish government pursued moderately expansionary fiscal policy throughout, for both financial rescue and, especially, countercyclical reasons. The EIU thought that Sweden's initial surplus and low debt were important reasons such expansion was feasible. The degree of stimulus was adjusted (in both directions) in response to the state of the economy.

Denmark, 2009:H1
Denmark initially pursued slightly expansionary fiscal policy for countercyclical reasons and financial rescue. But policy then turned generally moderately contractionary. The switch stemmed mainly from policymakers' ideas about the benefits of fiscal responsibility, with a secondary role for EU rules. There was only one minor mention of concern about market access.

Greece, 2009:H1
Greece's initial fiscal policy response consisted of a small expansion for financial rescue and small conventional fiscal actions in both directions driven by a range of motivations, with at most a small net fiscal expansion. But policy soon turned to austerity, which became increasingly severe over time. The initial turn to austerity stemmed from a combination of EU rules and issues with market access, with hints of a role for domestic policymakers' ideas. The later, harsher austerity came about as Greece lost market access and then turned to international organizations for aid, which came with strong conditionality.

Table 2. Summary of the Motivations for Fiscal Actions after Financial Crises, as Described in the Economist Intelligence Unit's *Country Reports*[a] (*Continued*)

Hungary, 2009:H1

Despite a small initial bank rescue, fiscal policy in Hungary was on net contractionary throughout, often strongly so. It was initially driven mainly by conditionality associated with an IMF-led program that predated the crisis. Later, it was driven mainly by concerns about market access, with some role for attempting to obtain renewed IMF support. A desire to comply with EU rules in order to join the euro area played a role at times, but was never central.

Ireland, 2009:H1

Ireland initially engaged in extreme financial rescue, partially counteracted by austerity in other areas. A loss of market access then led to extreme overall austerity. This was followed by continued austerity to comply with the conditionality of the IMF/EU bailout. Conditionality remained central through the end of the analysis period.

Sources: Economist Intelligence Unit, *Country Reports*; authors' analysis.

a. H = half year. See online appendix B for the detailed notes and quotations from the Economist Intelligence Unit's *Country Reports*, on which these summaries are based.

The top panel of table 3 divides the net fiscal expansions into those that were described by the EIU as small and those described as large. As can be seen, small expansions were more common than large ones. The bottom panel shows the motivations given by the EIU for the net fiscal expansion in each case. As discussed above, the three possible motivations—financial rescue, countercyclical stimulus, and politics—all correspond to policymakers' choices. Each line in the bottom panel corresponds to an episode, so that it is easy to see multiple motivations by episode.

FINANCIAL RESCUE One thing that jumps out from table 3 is that financial rescue is nearly universal. Of the 20 cases of deliberate net fiscal expansion, the EIU identified financial rescue as a motivation in 19 of them. The only case where it was not given as a motivation is the United States after its 1990 crisis; however, in this case a bailout of the savings-and-loan industry occurred before our narrative source identified substantial financial distress. Even in the two cases where there was no net fiscal expansion (Mexico and Hungary), the EIU reported substantial support for the financial system. This suggests remarkable agreement across policymakers from different countries that financial rescue is valuable and appropriate in times of high financial distress.

There is also an obvious correlation between the EIU's perceived size of the fiscal expansion and financial rescue. All 8 of the countries described as taking large postcrisis fiscal expansions are also described as being strongly motivated by financial rescue. For example, with respect to Finland after its 1993 crisis, the EIU wrote: "parliament approved a motion saying it

Table 3. Sizes of and Motivations for Fiscal Expansions in Episodes of High Financial Distress[a]

Size (date expansion is first mentioned is in parentheses)[b]	
Small	*Large*
United States (1992:Q1)[c]	Finland (1993:Q1)
Norway (1992:Q1)[d]	Sweden (1993:Q2)
South Korea (1999:Q1)[e]	Japan (1998:Q1)
Austria (2009:M1)	Turkey (2001:M7)
France (2009:M1)	United States (2008:M7)
Italy (2009:M1)[c]	Iceland (2009:M1)
Norway (2009:M1)	United Kingdom (2009:M1)
Portugal (2009:M1)	Ireland (2009:M7)
Spain (2009:M1)	
Sweden (2009:M1)[d]	
Denmark (2009:M7)	
Greece (2009:M7)[c]	

Motivation (date motivation is first mentioned is in parentheses)		
Financial rescue[f]	*Countercyclical*	*Politics*
	United States (1992:Q1)[g]	United States (1992:Q1)[g]
Norway (1992:Q1)	Norway (1992:Q1)	Norway (1993:Q1)[g]
Finland (1993:Q1)		
Sweden (1993:Q2)		
Japan (1998:Q1)	Japan (1998:Q1)	Japan (1998:Q1)[g]
South Korea (1999:Q1)[e]	South Korea (1999:Q1)	South Korea (1999:Q3)[g]
Turkey (2001:M7)		
United States (2009:M1)	United States (2008:M7)	
Iceland (2009:M1)		
United Kingdom (2009:M1)	United Kingdom (2009:M1)[g]	
Austria (2009:M1)	Austria (2009:M1)	
France (2009:M1)[g]	France (2009:M1)[g]	
Italy (2009:M1)[g]	Italy (2009:M1)[g]	
Norway (2009:M1)[g]	Norway (2009:M1)	
Portugal (2009:M1)	Portugal (2009:M1)	Portugal (2009:M1)
Spain (2009:M1)	Spain (2009:M1)	
Sweden (2009:M1)	Sweden (2009:M1)	
Denmark (2009:M7)	Denmark (2009:M7)	
Greece (2009:M7)[g]		
Ireland (2009:M7)		

Sources: Economist Intelligence Unit, *Country Reports*; authors' analysis.

a. Q = quarter; M = month. The results summarize the narrative evidence from the Economist Intelligence Unit's *Country Reports*. See online appendix B for the detailed quotations underlying our classifications.

b. Two countries had no net expansion after their crises (Mexico and Hungary).

c. Expansion was extremely small.

d. Expansion was toward the moderate direction.

e. South Korea undertook austerity first and then fiscal expansion.

f. In addition, both Mexico and Hungary, which did no net fiscal expansion, undertook substantial financial rescue (combined with other austerity measures).

g. Motivation was minor.

would grant sufficient funds and authorising the government to use them to secure ‘under all circumstances’ the continued operation of Finland’s banks” (EIU 1993:Q1, 8). Likewise, the EIU noted that in Iceland after the 2008 crisis, “gross government debt is forecast by the IMF to increase from 29% at the end of 2007 to 109% of GDP in 2009, as a result of meeting the obligations of the former three main Icelandic banks now taken into public ownership and the injection of new funds to recapitalise them” (EIU 2009, January, 9). That the largest fiscal expansions involved very aggressive financial rescue also suggests that policymakers viewed such rescue as a valuable use of public funds after a financial crisis.

COUNTERCYCLICAL STIMULUS A desire to stimulate aggregate demand and counter the contractionary consequences of a financial crisis is a less frequently cited motivation for postcrisis fiscal expansion. The EIU listed it as a motivation in 14 of the 20 cases of net expansion (and in 4 of those, it was only a minor motivation). This could suggest that the idea that fiscal stimulus is helpful and appropriate after a financial crisis was less widely held among policymakers than a belief in the efficacy and appropriateness of financial rescue.

The EIU’s descriptions provide interesting insight into the interaction between the financial rescue and countercyclical motivations. Some countries appear to have embraced both motivations strongly. For example, the EIU said of Japan after its 1998 crisis both that the “government . . . is now focusing on a bail-out of the financial sector” (EIU 1998: Q1, 3), and “in mid-November the government unveiled a new package of proposals designed to stimulate economic recovery” (EIU 1999:Q1, 17). After the 2008 crisis, the United States was similarly described as acting with gusto because of both motivations: “Mr. Obama has already made it clear that he will do everything to contain the crisis and that he is backing another fiscal stimulus package” (EIU 2009, January, 4).

Other countries appear to have embraced the financial rescue motivation strongly, but believed that they should undertake little or no conventional fiscal stimulus. For example, the EIU described the United Kingdom after the 2008 crisis as making “unprecedented moves last year to support the banking sector” (EIU 2009, January, 4), but only a “£20bn (US$30bn) fiscal stimulus package (the main element of which is a temporary cut in the rate of value-added tax (VAT) from 17.5% to 15% until end-2009)” (EIU 2009, January, 5). Indeed, at least three countries were described as believing in aggressive financial rescue and conventional fiscal austerity simultaneously (Sweden in the early 1990s, and Iceland and Ireland after the global financial crisis). For example, Ireland is described by the EIU as “taking all possible measures to support the financial system” (EIU 2009, July, 5), while, at nearly the same time, “[p]ublic spending is being cut as the government attempts to control an exploding budget deficit” (EIU 2010, January, 8).

Still other countries appear to have been relatively unmotivated by either financial rescue or countercyclical purposes, and therefore undertook only modest net fiscal expansion. This was the case with France and Italy after the 2008 crisis. For example, the EIU described the Italian government as taking “some modest anti-cyclical measures” (EIU 2009, January, 12), and having “guaranteed deposits up to about €100,000 and allowed banks to negotiate state help in recapitalisation, although no major Italian bank is at present in urgent need of state support” (p. 4).

POLITICS A final fact evident from the bottom panel of table 3 is that the EIU rarely attributed fiscal expansion after a financial crisis to political motives. One case where it did so was Portugal after the 2008 crisis. The EIU wrote: “The 2009 budget marked a break from the previous fiscal consolidation efforts, partly as a result of the economic slowdown, but also in light of the upcoming general election” (EIU 2009, January, 4). Even when political expediency was mentioned as a motivation, it was typically identified as a minor factor and was often seen as driving the precise timing of fiscal expansion rather than the overall direction. For example, in the case of Japan after its 1998 crisis, the EIU said: “With the possibility of a difficult lower house election and little sign of the hoped-for self-sustaining recovery in private-sector demand, the government will have little choice but to maintain a broadly accommodating fiscal policy in 2000–01” (EIU 2000:Q1, 8). That political considerations were rarely mentioned is perhaps not surprising, given that after financial crises there were often more proximate and obvious motivations for the EIU to discuss.

IV.C. Postcrisis Austerity

Table 4 summarizes the narrative evidence from the EIU on deliberate moves to net austerity during crisis episodes. The top panel shows that, according to the EIU, in 19 of the 22 episodes of high financial distress, governments eventually switched to net austerity. Most of these fiscal contractions were described as large, and typically occurred at least a year or two after the start of high distress. The only episodes where the EIU did not describe a deliberate move to net austerity were Japan after its 1998 crisis, and Sweden and Norway after the 2008 crisis.

The bottom panel of table 4 shows the motivations for austerity given by the EIU in each episode. As discussed above, the motivations reflect a mixture of market access constraints and policymakers' choices. Again, each line corresponds to an episode, so that multiple motivations are obvious.

MARKET ACCESS Problems with sovereign market access were cited as a motivation for fiscal contraction in 14 of the 19 moves to austerity. Although in some of these instances market access was a minor motivation or just one of several motivations, in at least half of them the EIU described market access as the primary or overarching motivation for austerity. This finding is somewhat at odds with the empirical evidence discussed in section III. The statistical analysis found some role for market access in driving the fiscal response to crises, but it was relatively minor and decidedly smaller than that of policymakers' choices. Overall, the EIU appeared to rank market access roughly on par with policymakers' choices in driving post-crisis austerity.

The EIU identified market access as a motivation for austerity in a wide range of situations. For example, it wrote of Sweden after its 1993 crisis: "The government is aware that the key is to reduce the budget deficit, as only then will it be able to obtain favourable credit conditions and reduce its public debt" (EIU 1995:Q3, 11). In this case, market access problems sound reasonably minor, and the government was undertaking austerity proactively to reduce interest rates on government bonds. The EIU described a more pressing market access motivation for Italy after the 2008 crisis: "The minister of the economy, Giulio Tremonti, appears determined to keep Italy's public finances under control, fearing that a severe deterioration would lead to a further sharp widening of interest rate spreads on Italy's government debt" (EIU 2009, January, 4). At the more extreme end of the spectrum is the case of Portugal after the 2008 crisis. The EIU wrote: "The Socialist Party (PS) government faces a major challenge to reduce the budget deficit, under severe pressure from financial markets" (EIU 2010, July, 3).

BAILOUT CONDITIONALITY When market access problems became severe, as they did in Portugal during the global financial crisis, countries were typically forced to turn to the IMF or other international bailout organizations. The EIU identified 8 cases where a country was forced to adopt austerity as a condition for international aid. For example, the EIU said of Portugal in 2011: "The bail-out . . . will depend on implementation of a severe fiscal squeeze, in order to reduce the government's budget deficit to below 3% of GDP and beyond" (EIU 2011, July, 5). Similarly, the EIU wrote of Ireland: "The government's economic policy will remain focused on implementing the austerity and financial and structural reforms agreed

Table 4. Sizes of and Motivations for Fiscal Austerity in Episodes of High Financial Distress[a]

	Size (date austerity is first mentioned is in parentheses)[b]	
	Small	*Large*
	United States (1991:Q1)[c]	Finland (1993:Q3)
	Norway (1994:Q3)	Sweden (1994:Q2)
	South Korea (1998:Q1)[d]	Mexico (1996:Q3)
	United States (2011:M7)	Turkey (2002:M1)
	France (2010:M7)[c]	Iceland (2009:M7)
	Denmark (2010:M7)	United Kingdom (2010:M7)
		Austria (2010:M7)
		Italy (2010:M7)[c]
		Portugal (2010:M7)
		Spain (2010:M1)
		Greece (2010:M1)
		Hungary (2009:M7)
		Ireland (2010:M1)

Motivation (date motivation is first mentioned is in parentheses)				
Market access	*Conditionality*	*Ideas*	*EU rules*	*Countercyclical*
		United States (1991:Q1)		
		Norway (1994:Q3)		Norway (1994:Q3)[e]
Finland (1994:Q3)[e]		Finland (1993:Q3)	Finland (1995:Q1)	
Sweden (1995:Q1)		Sweden (1994:Q2)	Sweden (1994:Q3)	
Mexico (1996:Q3)	Mexico (1996:Q3)			
South Korea (1998:Q1)	South Korea (1998:Q1)			
Turkey (2002:M1)	Turkey (2002:M1)		Turkey (2005:M1)	
		United States (2011:M7)		
Iceland (2010:M1)	Iceland (2009:M7)	Iceland (2010:M1)[e]		
United Kingdom (2010:M7)[e]			United Kingdom (2010:M7)	
		Austria (2010:M7)	Austria (2010:M7)	
France (2010:M7)		France (2010:M7)	France (2010:M7)	
Italy (2010:M7)			Italy (2011:M1)[e]	
Portugal (2010:M7)	Portugal (2011:M7)	Portugal (2012:M1)[e]	Portugal (2010:M7)	
Spain (2010:M1)		Spain (2010:M1)[e]	Spain (2011:M1)[e]	
		Denmark (2010:M7)	Denmark (2011:M1)	Denmark (2012:M7)[e]
Greece (2010:M1)	Greece (2010:M7)		Greece (2010:M1)	
Hungary (2009:M7)	Hungary (2009:M7)	Hungary (2009:M7)[e]	Hungary (2009:M7)	
Ireland (2010:M1)	Ireland (2011:M1)			

Sources: Economist Intelligence Unit, *Country Reports*; authors' analysis.

a. Q = quarter; M = month. The results summarize the narrative evidence from the Economist Intelligence Unit's *Country Reports*. See online appendix B for the detailed quotations underlying our classifications.

b. Three countries never on net undertook austerity after their crises (Japan, Sweden 2008, and Norway 2008).

c. Toward the moderate direction.

d. South Korea undertook austerity first and then fiscal expansion.

e. Motivation was minor.

in exchange for access to a €85bn EU/IMF lending facility" (EIU 2011, January, 3).

As can be seen from table 4, the cases where market access problems were severe enough to lead to an IMF rescue typically involved large net fiscal contractions. For example, the EIU wrote of Greece: "In May 2010 Greece signed a Memorandum of Understanding (MoU) with representatives of the European Commission, European Central Bank (ECB), and IMF (the 'troika'), which committed it to a draconian programme of fiscal consolidation and economic reform" (EIU 2011, July, 5). The one exception to this pattern was South Korea after its 1998 crisis. South Korea was described by the EIU as following required austerity only for a relatively short time. It is perhaps not a coincidence that the EIU reports also contained frequent mentions of the fact that "South Korea's public finances are sound compared with those of many other OECD countries. Gross government debt stands at around 10% of GDP" (EIU 2011, August, 11).

IDEAS Although the EIU clearly believed that market access problems and bailout conditionality were motivations for postcrisis austerity in a number of cases, it suggested that policymakers' choices also played an important role. The EIU identified domestic policymakers' ideas as a motivation for austerity in 13 of the 19 cases of deliberate postcrisis fiscal contraction.

In some cases, the EIU was quite explicit about the role of ideas in motivating austerity. For example, according to the EIU, Finland adopted austerity after its 1993 crisis because the "centre-right government" believed it would "reduce unemployment by stimulating the private sector" (EIU 1993:Q3, 9). Later in this episode, the EIU confirmed the role of ideas when it reported that the president "has been quick to reassert that the problem [of high unemployment] cannot be combated by more state borrowing" (EIU 1994:Q3, 15). Likewise, after engaging in both financial rescue and fiscal stimulus in the immediate aftermath of the 2008 crisis, policymakers in Denmark switched to austerity in mid-2010. According to the EIU, the prime minister "used his final New Year address before the next general election to highlight the need for deeper structural reform of the welfare system to prepare Denmark for the future fiscal challenges of population ageing" (EIU 2011, January, 13).

The EIU's descriptions of the United Kingdom's move to austerity in mid-2010 provide a somewhat more circumstantial case that ideas were a key motivation. The EIU reported that "the chancellor justified the extent of the fiscal squeeze on the grounds that it was needed in order to retain the confidence of the markets." It then went on to say, however, that "[d]espite record issuance, the UK bond market has been one of the strongest in industrialised economies this year" (EIU 2010, July, 16). The juxtaposition of the stated fear of market access problems and benign actual conditions suggests that ideas about the harm caused by deficits and high debt were key. The EIU's statement that "the Conservative chancellor of the exchequer, George Osborne, insists that weak economic activity will not deflect the coalition from its aggressive deficit-reduction plans" (EIU 2011, July, 13) could also suggest an ideological motivation for the austerity.

The case of the United Kingdom makes it clear that even large fiscal contractions after a financial crisis can be motivated by ideas. The EIU said of British policy in 2010: "The scale of fiscal consolidation implied in the budget is immense. Total discretionary tightening of £113bn (US$170bn) a year is planned by 2014/15 (April–March), compared with current levels, equivalent to 6.5% of projected GDP" (EIU 2010, July, 6). Austria is another case where substantial austerity was driven by ideas—and EU rules. The EIU repeatedly described contractionary moves of 2 percent of GDP or more despite no issues with market access (EIU 2010, July; 2012, July).

Unfortunately, the EIU did not provide much information about the precise ideas that motivated policymakers' choices to adopt austerity. One idea that it sometimes cited was concern about long-run market access. For example, in discussing the switch to austerity in the United States in 2011, it reported, "Although there is no immediate pressure on the government's finances, Washington appears set on tackling the deficit aggressively" (EIU 2011, July, 4). It also said, "The government faces no funding pressures at present . . . . However, . . . [h]igh debt levels create the risk of an eventual rise in US bond yields that would increase borrowing costs" (p. 6). Another example is provided by Sweden's decision to pursue contraction in terms of conventional fiscal policy after its banking crisis in the early 1990s. In the context of a discussion of falling long-term interest rates, the need to attract foreign capital, and the government's contractionary fiscal policy, the EIU said, "The central government wants to keep up this level of enthusiasm. Luckily, its credit rating is still good and should remain so" (EIU 1993:Q3, 12). The United Kingdom's shift to austerity in 2010 also appears to largely fit into this category. As described above, the government cited the need to maintain market access as a key motivation for austerity despite exceptionally low long-term interest rates. Indeed, the only potentially imminent issue with market access the EIU reported in this entire episode was a possible downgrade of the United Kingdom's sovereign bond rating from AAA (EIU 2009, July, 6; 2010, January, 3).

Thus, the concerns involving market access appear to have been largely long-term.[34]

Concerns about future market access are related to the idea that a financial crisis can lead to a sovereign debt crisis (for example, Reinhart and Rogoff 2009). After the global financial crisis, two other ideas that received considerable attention are that crossing a specific debt threshold can greatly harm growth (an idea that was based in part on a simplified reading of Reinhart and Rogoff 2010), and that austerity can be expansionary, particularly if it focuses on spending cuts (for example, Alesina and Ardagna 2010). However, these ideas barely registered in the EIU's *Country Reports*. The closest the EIU came to discussing a debt threshold came in its analysis of France, where it often mentioned the possibility that debt would reach 90 percent of GDP (which was a level that featured prominently in Reinhart and Rogoff 2010). Its strongest statement was, "With public debt forecast to rise to close to 90% of GDP by 2012, an additional risk is that France could lose its AAA rating on sovereign debt, which would push up interest costs" (EIU 2011, January, 8). But this idea did not appear in other episodes. And even in the case of France, the EIU did not explicitly state that it attached particular importance to the 90 percent figure, rather than just using it as a convenient round number to describe the trajectory of France's debt.

Similarly, ideas related to the work of Alberto Alesina and Silvia Ardagna (2010) received only a few passing mentions. For example, at one point the EIU attributed the composition of fiscal policy in Ireland, but not its overall direction, to these ideas. In explaining why a budget consolidation was "to be achieved almost exclusively by expenditure reductions," it said, "the government has explicitly acknowledged that international evidence points to spending cuts being a more effective route to consolidation than tax increases" (EIU 2010, January, 5–6). Similarly, the EIU implicitly cited this line of work in giving its own views of fiscal policy in the United Kingdom, but it did not attribute this view to policymakers. It said, "The government has been criticised in some quarters for targeting too rapid a pace of deficit reduction at a time of economic uncertainty and minimal funding pressures, but there is no

34. The one piece of evidence in the other direction is that the EIU reported, "The budget's austere tone has reduced near-term market concerns over fiscal sustainability" (July 2010, 6). However, because this discussion came soon after the fears of the possible downgrade, and because the EIU also reported that "there is no indication that investors are shunning U.K. gilts" (p. 16), it appears that the EIU was referring to the potential downgrade, not the possibility of severe loss of market access.

accepted orthodoxy on the impact that fiscal austerity will have on economic activity" (EIU 2011, January, 7).

Most often, however, the EIU did not provide specific information about the ideas that motivated policymakers' choices to pursue austerity. Instead, it simply reported that policymakers viewed high debt or deficits as problems that needed to be addressed. For example, in discussing Norway's shift to austerity in the wake of its banking crisis in the early 1990s, it said, "policymakers are having to concentrate on potential long-term problems. One of these is the budget deficit" (EIU 1994:Q3, 6). Similarly, in discussing Denmark's move to austerity after the global financial crisis, the EIU referred to "[t]he need for budgetary consolidation," without explaining the reasons for the need (July 2010, 5). And, it reported that "the deterioration in the public finances has also made the electorate aware of the long-term costs of sustaining Denmark's welfare benefits" (p. 4). Another example comes from the EIU's discussion of France's turn to austerity in the wake of its crisis in 2008. It said simply that one "legacy of the 2008–09 downturn" is "a need for fiscal tightening," without explaining why (EIU 2011, July, 8).

EU FISCAL RULES A motivation for austerity after financial distress also related to ideas involved the fiscal rules set by the European Union. As discussed above, such rules can be thought of as reflecting the ideas of EU policymakers. This is especially true because the EU appears to have had substantial leeway in how aggressively it pressured member countries to conform to its guidelines. There was also a role for domestic policymakers' ideas in how willingly they acceded to EU wishes. This was especially true in cases where countries worked to meet the guidelines out of a desire to qualify for membership.

The EIU mentioned EU fiscal rules as a motivation for fiscal actions in 11 of the 19 cases of postcrisis austerity in our sample. Four of these cases involved countries taking actions in advance of participation in an EU program. The EIU wrote of Sweden in 1995: "The government is bullish about the fiscal outlook. Its EMU [Economic and Monetary Union] convergence plan presented in June includes new savings measures to be introduced, if needed, from 1997" (EIU 1995:Q3, 1). Finland after its 1993 crisis was also described as undertaking austerity to satisfy EMU criteria (EIU 1995:Q3, 5). Likewise, both Turkey's and Hungary's postcrisis austerity was partly attributed to desire to join the EU. For example, the EIU said of Turkey: "The IMF will provide the extra lending in return for government abidance by a programme of tight fiscal policy backed by privatisation and structural reforms . . . . This programme had already been set out in the three-year Pre-Accession Economic Programme (PEP) drawn up for the EU on November 30th" (EIU 2005, January, 19).

The other cases where the EIU mentioned EU rules as a motivation for postcrisis austerity involve existing EU members. For example, it wrote of Austria after the 2008 crisis: "The coalition has agreed in principle with the European Commission in its 2010–13 Stability Programme to act to reduce the government deficit to below 3% of GDP by 2013" (EIU 2011, July, 5). Similarly, for France after the global financial crisis, the EIU wrote that "Germany will press for more intrusive budget monitoring of euro area members, which would be likely to stoke further tension with France, given the country's poor fiscal record and habit of flouting the euro area's fiscal rules when these conflict with its domestic priorities" (EIU 2010, July, 5). This statement makes it clear that EU rules and actions sometimes reflected the ideas of foreign policymakers.

COUNTERCYCLICAL AUSTERITY Very rarely, the EIU described countries as undertaking austerity at some point after a crisis for countercyclical purposes. That is, policymakers tightened fiscal policy to prevent the economy from overheating. The most obvious case involved Norway after its crisis in the early 1990s. As can be seen from the top panel of figure 3, Norway grew much more rapidly after its crisis than one would have predicted based on its previous history. The EIU wrote: "as growth accelerates the government will continue to introduce spending cuts in a counter-cyclical fashion" (EIU 1995:Q1, 5), and, "as growth is picking up it finds itself in a position to make the necessary cuts, while at the same time overseeing a reduction, if not the eradication, of the budget deficit" (p. 11).

Overall, the EIU's discussion of the motivation for postcrisis austerity paints a mixed picture. The EIU clearly saw an important role for market access problems, often accompanied by an IMF bailout and conditionality, in driving austerity. At the same time, domestic policymakers' ideas, often interacting with EU rules and ideas, were also seen as an important factor. Thus, the EIU provides narrative evidence that postcrisis austerity was often at least partly a choice.

IV.D. Narrative Evidence on the Role of Debt in Driving the Fiscal Response to Crises

A central finding from our statistical work is that the debt-to-GDP ratio has substantial predictive power for the fiscal policy response to financial distress, and that this is true even when we control for more direct measures of market access. It therefore makes sense to ask what our narrative analysis shows about the link between debt and the fiscal policy response to a crisis.

In examining the EIU's *Country Reports* for evidence about this issue, we look for cases where the EIU explicitly cited countries' debt as mattering, as opposed to more general comments about influences from countries' fiscal situations. And we focus on cases where the EIU saw debt as affecting the views and behavior of investors, policymakers, and international organizations, rather than ones where it appeared to be just expressing its own concerns. Of course, the EIU probably did not report every instance where debt influenced the policy response. But we take discussion of the role of debt in the *Country Reports* as an indication that debt was important.

The first finding is that the EIU often described debt as affecting the fiscal policy response to a crisis. In 18 of our 22 episodes of high distress, the EIU believed that debt had at least some impact on the policy response.[35]

In two of those episodes (Norway and Sweden after the 2008 crisis), the EIU saw low debt as an important factor behind fiscal expansion. For example, in explaining Sweden's continued expansionary policy well after the peak of its financial distress, the EIU said: "With public debt and the government deficit low, the coalition has been able to introduce some stimulus measures" (EIU 2012, January, 3).

In the remaining 16 cases, the EIU described high or rising debt as a force fostering fiscal austerity or limiting expansion. The information about the channels through which the EIU described debt as affecting policy in these episodes is summarized in table 5. The most common channel was through the views of domestic policymakers. The level or growth of debt was frequently invoked as a reason to limit expansion or switch to contraction, and altering the path of the debt-to-GDP ratio was often an explicit goal of policy. For example, the EIU reported that a contractionary budget in Finland after its 1993 crisis was "intended as part of a drive to ensure that state debt does not rise beyond 70% of GDP"

35. The four episodes where the EIU did not describe a link are Norway (1991), Mexico, South Korea, and Ireland. The EIU's discussion of policy in Norway following its crisis included one reference to the possibility of net debt becoming positive, and of this development having the potential to affect policy (EIU 1992:Q3, 5). But the EIU did not attribute this view to policymakers or tie it to actual policy. In the cases of Mexico and South Korea, the EIU reported severe problems with market access, but it did not connect them with the level or trajectory of debt. And in the case of Ireland, the EIU viewed the large *deficits* arising from its massive financial rescue as an important source of its adoption of austerity, but again it did not assign a role to debt.

Table 5. Narrative Evidence on the Role of Debt in Fostering Austerity or Limiting Expansion in Episodes of High Financial Distress[a]

Via market access	*Via conditionality*	*Via ideas*	*Via EU rules*
		United States (1991:Q3)[b]	
		Finland (1993:Q3)	Finland (1995:Q1)
Sweden (1996:Q3)		Sweden (1994:Q3)	Sweden (1996:Q3)
		Japan (2000:M6)	
Turkey (2003:M7)	Turkey (2002:M1)		
		United States (2011:M7)[b]	
		Iceland (2009:M1)[b]	
		United Kingdom (2010:M7)	
		Austria (2010:M1)	
France (2011:M1)		France (2010:M1)[b]	
Italy (2009:M7)			Italy (2011:M1)
	Portugal (2011:M7)		
Spain (2010:M7)			
		Denmark (2009:M7)[b]	
Greece (2009:M7)	Greece (2010:M7)		
Hungary (2010:M1)	Hungary (2011:M7)		

Sources: Economist Intelligence Unit, *Country Reports*; authors' analysis.

a. Q = quarter; M = month. The results summarize the narrative evidence from the Economist Intelligence Unit's *Country Reports* about episodes where debt was described as affecting fiscal policy through various channels. Date motivation is first mentioned is in parentheses. See online appendix B for the detailed quotations underlying our classifications.

b. Weaker evidence.

(EIU 1993:Q3, 4). Similarly, the EIU described how, soon after Austria's 2008 crisis, the vice chancellor "indicated that the government will begin to unwind its fiscal stimulus measures as it seeks to prepare for austerity measures" (EIU 2010, January, 5), and that he "rejected a third stimulus package, claiming that Austria must focus on reining in its rising government deficit and public debt" (p. 11). In some of the cases, the apparent link between debt and idea-driven policy is more tenuous. For example, in analyzing the U.S. response to its 2007 crisis, the EIU reported: "the rise in federal debt . . . is . . . a serious concern" (EIU 2010, January, 6), and, "US public debt has risen sharply in recent years . . . . Washington appears set on tackling the deficit aggressively" (EIU 2011, July, 4). But it did not clearly tie these views to specific policy actions.

Table 5 also shows that in a few cases, direct concern about the level of debt had substantial effects through EU fiscal rules. In describing fiscal policy in Finland a few years after its crisis, for example, the EIU said: "Economic policy will be driven by the government's determination to take part in the third stage of Economic and Monetary Union . . . . The most immediate threat to Finland's qualifying for inclusion is its public debt . . . . Fiscal policy will therefore remain tight" (EIU 1995:Q3, 5).

Importantly, debt also played a significant role through its influence on market access. In many cases, the link was direct: debt was a source of concern to investors, and thus led to problems with market access. For example, the EIU reported that in Italy after the global financial crisis: "The spread of the Greek debt crisis since the end of 2009 forced Italy . . . to reduce [its] budget deficits and government debt burdens in order to prevent a further loss of investor confidence" (EIU 2010, July, 6). As another example, in describing Turkish fiscal policy in the wake of its 2001 crisis, the EIU linked the government's austerity to a desire to avoid a sovereign debt crisis, and it linked the possibility of such a crisis to investors' concerns about Turkey's high debt ratio: "A tighter 2003 budget than initially expected [and other factors] . . . should be enough for Turkey to avoid a crisis this year. But because of the sheer size of Turkey's government debt (about 85% of GDP at end-2002), the widening current-account deficit and Turkey's vulnerability to sudden shifts of investor sentiment, the risk of a financial crisis will remain high during the outlook period" (EIU 2003, July, 9).

In several cases, the EIU discussed loss of market access leading countries to turn for help to international organizations, which then imposed conditionality that put considerable emphasis on the behavior of debt. One example of such conditionality is provided by Portugal after its crisis.

The EIU stated: "A bail-out from the joint EU/IMF emergency credit facility will entail strict fiscal consolidation" (EIU 2011, July,3), a key goal of which would be "to stabilise public debt by 2013" (p. 6). Another example comes from Hungary after its crisis, where the EIU reported: "The [European] Commission and the IMF both called for an exact timetable on Hungary's path towards a state debt ceiling of 50% of GDP" (July 2011, 13).

A final finding, which is not shown in table 5, concerns the *type* of fiscal policy that the EIU viewed as being influenced by debt: In every case where the EIU perceived debt as promoting austerity or limiting expansion, it was through its influence on conventional fiscal policy. That is, there were no cases where the EIU reported that government debt limited the extent of financial rescue. The closest was a hypothetical about Italy: "In the event of a collapse of the banking sector . . . , the Italian state would have less credibility than Germany, France, or the UK if it were to intervene in a dramatic way, since public debt is over 100% of GDP" (EIU 2009, January, 4). But even in this case, the EIU did not view debt as limiting financial rescue. This reinforces the evidence for near-universal support for rescue in the event of severe financial distress.

The strong evidence from the narrative analysis that debt was often a driver of the fiscal policy response to a crisis is consistent with—and helps explain—the finding from our statistical work that the debt-to-GDP ratio is a powerful predictor of the policy response, even controlling for what are arguably more direct measures of market access. The narrative analysis finds that not just investors, but also domestic policymakers, the EU, and the IMF, used the level of debt and its changes as important markers of countries' fiscal health and as an important guide to policy. Thus, it is not surprising that we find that the debt-to-GDP ratio appears to be an important determinant of how countries respond to financial distress, and that it appears to matter even when we control for direct measures of market access.

IV.E. The EIU's Assessments of the Fiscal Response to a Crisis and Data on the Prior Debt Ratio

Our final exercise is to examine the relationship between the EIU's descriptions of the fiscal response to financial distress and countries' prior debt-to-GDP ratios. Specifically, for each of the various groups of episodes reported in tables 3, 4, and 5, we find the average debt ratio across the episodes in the group. In this way, we are blending the narrative evidence on the sizes and motivations of the fiscal response to crises with actual data on the debt-to-GDP ratio. As in our statistical work, to avoid direct reverse causation from fiscal policy to debt, we use figures for debt as of the end of the year before when the EIU first described a development.[36] If debt is an important determinant of how fiscal policymakers respond to financial distress, one would expect to see a connection between the prior debt-to-GDP ratio and the EIU's accounts of the nature and determinants of fiscal policy after crises. At the same time, because the number of observations in each category is small and the EIU's general descriptions of fiscal policy are presumably less precise than our data, we do not view this evidence as central to our analysis.

Table 6 presents the results. The patterns go in the directions one would expect, but not overwhelmingly so. Debt was on average slightly lower in situations where the EIU saw the adoption of highly expansionary policy after high financial distress than in cases where it saw a move to only small (or no) expansion; and debt was noticeably higher when the EIU reported the adoption of sharp austerity than when it reported a shift to only mild austerity.[37] The debt ratio was typically 10 to 15 percentage points higher in situations where the EIU first reported a given motivation for austerity than when it first saw a given motivation for expansion; this is what one would expect if debt affects the direction of the policy response. Also, consistent with our findings that debt affects the policy response to financial distress through multiple channels, the average debt ratio varied little across cases with different motivations for expansion, or across cases with different motivations for austerity.

The results are stronger when we turn to cases where the EIU reported an impact of debt on policy—perhaps not surprisingly, the EIU perceived debt as having a more contractionary influence on policy when the debt-to-GDP ratio was larger. Most notably, debt was much lower in the cases where the EIU reported that debt fostered expansionary policy, or when it never described an influence of debt, than in the cases where it saw debt

36. Recall that we use the first EIU report of a year largely to gain insight into developments late in the year before. For developments first described in these reports, we therefore use the debt number from the end of the calendar year two years before that of the report. Thus, for example, for the U.S. financial rescue, which was first reported in the January 2009 report, we use the debt number for 2007. When we report debt numbers for cases where the EIU did not describe a development (such as episodes where the EIU never saw austerity), we use the figures from the year before distress first reached 7. Finally, because our debt data for Mexico do not begin until 1996, the figures we report do not include Mexico's crisis episode.

37. As an accounting matter, the reason the average debt ratio is quite high for the three countries with no austerity is that one of them —Japan—had very high debt.

Table 6. The Relationship between the EIU's Assessments of the Fiscal Response to a Crisis and Data on the Prior Debt Ratio[a]

EIU description	*Average debt ratio (percent)*
Size of expansion	
None	71 (1)
Small	55 (12)
Large	49 (8)
Motivation for expansion	
Financial rescue	52 (19)
Countercyclical	55 (14)
Politics	57 (5)
Size of austerity	
None	63 (3)
Small	57 (6)
Large	68 (12)
Motivation for austerity	
Market access	66 (13)
Conditionality	68 (7)
Domestic ideas	65 (13)
EU rules	74 (11)
Countercyclical	49 (2)
Role of debt	
No role mentioned	26 (3)
Fostered expansion	44 (2)
Fostered austerity	
Via market access	80 (7)
Via conditionality	87 (4)
Via domestic ideas	66 (10)
Via EU rules	78 (3)

Source: Authors' calculations.

a. Debt figures are based on when developments were first mentioned by the EIU in an episode, using numbers for the previous year. The number of observations is given in parentheses.

contributing to austerity.[38] And looking across the various channels through which the EIU reported that debt affected austerity, the average debt ratio was relatively high for all of them. It was highest when debt affected austerity through conditionality and market access, and it was lowest when its effects operated through the ideas of domestic policymakers. Thus, the relationship between the EIU's descriptions of the fiscal policy response to a crisis and the actual debt ratio provides additional support for the findings that debt affects the policy response to financial distress, and that it does so both via market access and through other mechanisms.

38. The reason the entry for "No role mentioned" is based on three observations rather than four is that, as described above, we do not have debt data for Mexico's crisis episode.

V. Conclusions

This paper examines the fiscal policy response to financial crises, focusing especially on the question of why some countries use fiscal policy so much more aggressively in the face of a crisis than others. Here, we summarize out findings and discuss their potential implications for policy.

V.A. Summary

There is compelling evidence that a country's fiscal response to a crisis is correlated with its prior debt-to-GDP ratio. Countries that face a crisis with lower debt pursue more expansionary fiscal policy than countries that face a crisis with higher debt. Indeed, our estimates based on data for 30 OECD countries since 1980 suggest that facing a crisis with a debt ratio 1 standard deviation below the sample average is associated with a high-employment surplus in the aftermath of a crisis that is roughly 4 percentage points smaller (that is, more expansionary) than facing a crisis with a debt ratio 1 standard deviation above the average. Moreover, this large difference in the policy response is associated with a large difference in the aftermath: low-debt countries experience much milder downturns after a crisis than high-debt countries.

This paper seeks to understand why the prior debt-to-GDP ratio appears to matter for the fiscal response to crises. We find that including direct measures of sovereign market access, such as long-term government bond rates or sovereign credit ratings, does not eliminate, or even greatly attenuate, the estimated impact of the debt-to-GDP ratio on the high-employment surplus after crises. This strongly suggests that the debt ratio does not matter simply through its impact on current market access or because it is a proxy for market access. Rather, an important part of why it matters appears to be through its impact on policymakers' choices.

Narrative evidence from the Economist Intelligence Unit suggests a somewhat larger role for market access in driving the fiscal response to crises than the statistical evidence. This is especially true if IMF fiscal conditionality is thought of as an extension of market access problems. But the evidence from the EIU also indicates that policymakers' choices were often important. In at least half the cases of postcrisis austerity, the EIU assigned a central role to the ideas of domestic policymakers or the European Union's fiscal rules. Moreover, the EIU provides evidence that policymakers' choices depend in substantial part on the debt-to-GDP ratio. Thus, it confirms that the debt ratio matters in considerable part through those choices.

Taken together, the statistical and narrative evidence suggests that both sovereign market access and policymakers' choices help account for the observed correlation between a country's debt-to-GDP ratio and the fiscal response to crises. The combined weight of these two types of evidence, however, points to policymakers' choices as somewhat more central.

V.B. Policy Implications

There are two directions one might be tempted to go in drawing policy implications from these findings. The first would be to argue that the fact that a high debt-to-GDP ratio is clearly associated with worse postcrisis outcomes is a compelling reason for countries to keep their debt ratio in check. By doing so, a country reduces the chance that it will face severe market access problems that force it to respond to a crisis with austerity. It also means that domestic policymakers are less likely to choose counterproductive austerity after a crisis or face pressure from external organizations to do so.

But our finding that the debt ratio matters for the postcrisis fiscal response in substantial part through policymakers' choices suggests a second possibility. Policymakers in countries with good market access could choose not to base postcrisis fiscal policy on the debt ratio. That is, rather than respond to the evidence that debt matters for the fiscal response by reducing debt proactively, policymakers could change their ideas about the desirable fiscal response to a financial crisis when the debt ratio is high. Similarly, the EU could adopt a more flexible set of fiscal rules that would allow member countries to expand aggressively after a crisis, regardless of debt levels. Even the IMF, which is only called in when market access disappears, could adopt less rigid fiscal constraints as a condition for its assistance.

Because we find that debt matters for the fiscal response to a crisis through both market access and policymakers' choices, surely the most sensible policy prescription is not to view these two possibilities as competing but to strongly embrace both of them. Countries should work to keep debt low as an insurance policy for future crises and to minimize market risks. But, when confronted with high financial distress, domestic policymakers and leaders of international organizations should not let debt loads unnecessarily drive the fiscal response. To do so leads to much worse postcrisis output losses.

Such a combined strategy obviously involves important dynamic considerations. If high-debt countries do not choose tight fiscal policy after a crisis, it is possible that market access problems will eventually develop.

But this possibility does not argue for the status quo of high-debt countries with ample market access choosing austerity after a crisis. Rather, it argues for a new consensus among policymakers that stresses running responsible fiscal policy in ordinary times, and undertaking aggressive fiscal expansion if at all possible in response to a financial crisis.

ACKNOWLEDGMENTS We are grateful to James Church for assistance with data; to Olivier Blanchard, Janice Eberly, Patrick Honohan, Lars Jonung, Philip Lane, Edward Nelson, Maurice Obstfeld, James Stock, Phillip Swagel, Brookings Panel participants, and seminar participants at the University of California, Berkeley, for helpful comments; and to the University of Edinburgh for support during early stages of this project through the U.K. Economic and Social Research Council, award reference ES/L990633/1.

References

Alesina, Alberto, and Silvia Ardagna. 2010. "Large Changes in Fiscal Policy: Taxes versus Spending." *Tax Policy and the Economy* 24, no. 1: 35–68.

Baron, Matthew, Emil Verner, and Wei Xiong. 2019. "Salient Crises, Quiet Crises." Working paper. wxiong.mycpanel.princeton.edu/papers/BankingCrisis.pdf.

Bernanke, Ben S. 1983. "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression." *American Economic Review* 73, no. 3: 257–76.

Bernardini, Marco, and Lorenzo Forni. 2018. "Private and Public Debt Interlinkages in Bad Times." Working paper. https://lorenzoforni.files.wordpress.com/2018/07/private-and-public-debt-interlinkages-in-bad-times.pdf.

Binder, Carola Conces. 2018. "Political Pressure on Central Banks." Working paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3244148.

Bohn, Henning. 1998. "The Behavior of U.S. Public Debt and Deficits." *Quarterly Journal of Economics*, 113, no. 3, 949–63.

Bordo, Michael, Barry Eichengreen, Daniela Klingebiel, and Maria Soledad Martinez-Peria. 2001. "Is the Crisis Problem Growing More Severe?" *Economic Policy* 16, no. 32: 51–82.

Chodorow-Reich, Gabriel. 2019. "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" *American Economic Journal: Economic Policy* 11, no. 2: 1–34.

Chodorow-Reich, Gabriel, Laura Feiveson, Zachary Liscow, and William Gui Woolston. 2012. "Does State Fiscal Relief during Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act." *American Economic Journal: Economic Policy* 4, no. 3: 118–45.

EIU (Economist Intelligence Unit). Various years. *Country Reports*. London: Economist Intelligence Unit Limited.

Elmendorf, Douglas. 2016. "Recommendations for Federal Fiscal Policy." Speech delivered at Federal Reserve Bank of San Francisco. https://www.hks.harvard.edu/more/about-us/leadership-administration/offices-deans/deans-office/presentations-and-papers/recommendations-for-federal-fiscal-policy.

Fisher, Jonas D. M., and Ryan Peters. 2010. "Using Stock Returns to Identify Government Spending Shocks." *Economic Journal* 120, no. 544: 414–36.

Ghosh, Atish R., Jun I. Kim, Enrique G. Mendoza, Jonathan D. Ostry, and Mahvash S. Qureshi. 2013. "Fiscal Fatigue, Fiscal Space and Debt Sustainability in Advanced Economies." *Economic Journal* 123, no. 566: F4–F30.

Girouard, Nathalie, and Christophe André. 2005. "Measuring Cyclically Adjusted Budget Balances for OECD Countries." Working Paper 434. Paris: Organization for Economic Cooperation and Development. https://doi.org/10.1787/787626008442.

Guajardo, Jaime, Daniel Leigh, and Andrea Pescatori. 2014. "Expansionary Austerity? International Evidence." *Journal of the European Economic Association* 12, no. 4: 949–68.

Ilzetzki, Ethan, Enrique G. Mendoza, and Carlos A. Végh. 2013. "How Big (Small?) Are Fiscal Multipliers?" *Journal of Monetary Economics* 60, no. 2: 239–54.

IMF (International Monetary Fund). 2018. *IFS World and Country Notes, Yearbook 2018*. Washington: International Monetary Fund.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2016. "Sovereigns versus Banks: Credit, Crises, and Consequences." *Journal of the European Economic Association* 14, no. 1: 45–79.

Kose, M. Ayhan, Sergio Kurlat, Franziska Ohnsorge, and Naotaka Sugawara. 2017. "A Cross-Country Database of Fiscal Space." Development Prospects Group Working Paper 8157. Washington: World Bank. http://documents.worldbank.org/curated/en/601211501678994591/pdf/WPS8157.pdf.

Laeven, Luc, and Fabián Valencia. 2014. "Systemic Banking Crises." In *Financial Crises: Causes, Consequences, and Policy Responses*, edited by Stijn Claessens, M. Ayhan Kose, Luc Laeven, and Fabián Valencia. Washington: International Monetary Fund.

Mendoza, Enrique G., and Jonathan D. Ostry. 2008. "International Evidence on Fiscal Solvency: Is Fiscal Policy 'Responsible'?" *Journal of Monetary Economics* 55, no. 6: 1081–93.

Nakamura, Emi, and Jón Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from U.S. Regions." *American Economic Review* 104, no. 3: 753–92.

Obstfeld, Maurice. 2013. "On Keeping Your Powder Dry: Fiscal Foundations of Financial and Price Stability." *Monetary and Economic Studies* 31: 25–38.

OECD (Organization for Economic Cooperation and Development). Various years. *OECD Economic Outlook*. Paris: OECD Publishing.

Perotti, Roberto. 1999. "Fiscal Policy in Good Times and Bad." *Quarterly Journal of Economics* 114, no. 4: 1399–436.

Ramey, Valerie A. 2011. "Identifying Government Spending Shocks: It's All in the Timing." *Quarterly Journal of Economics* 126, no. 1: 1–50.

———. 2016. "Macroeconomic Shocks and Their Propagation." In *Handbook of Macroeconomics, Volume 2*, edited by John B. Taylor and Harald Uhlig. Amsterdam: Elsevier.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly*. Princeton, N.J.: Princeton University Press.

———. 2010. "Growth in a Time of Debt." *American Economic Review* 100, no. 2: 573–78.

Romer, Christina D., and David H. Romer. 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review* 100, no. 3: 763–801.

———. 2017. "New Evidence on the Aftermath of Financial Crises in Advanced Economies." *American Economic Review* 107, no. 10: 3072–118.

———. 2018. "Why Some Times Are Different: Macroeconomic Policy and the Aftermath of Financial Crises." *Economica* 85, no. 337: 1–40.

Suárez Serrato, Juan Carlos, and Philippe Wingender. 2016. "Estimating Local Fiscal Multipliers." Working Paper 22425. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w22425.

Comments & Discussion

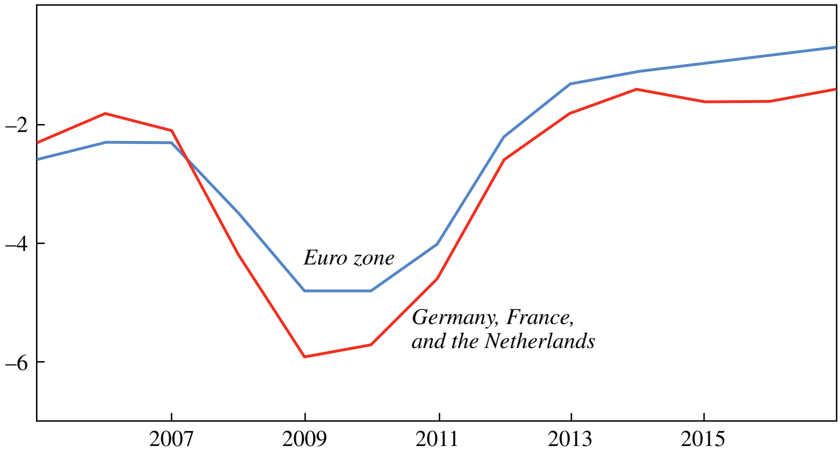
COMMENT BY

OLIVIER BLANCHARD This is a great paper, in the Romer-and-Romer style, which I see as the emerging gold standard for empirical macroeconomics. It painstakingly constructs a new data set (I assumed this involved heavy undergraduate inputs or machine-learning algorithms. I am told, however, by the authors that they did it all themselves, a questionable accomplishment in the light of comparative advantage). Their paper uses both qualitative and quantitative information, marrying econometrics and case studies, with informal and formal analyses of the consistency between the two. It is honest and, as a result, it is credible; when the results do not quite fit the priors, they are not swept under the rug. And, most important, they look at what I see, admittedly through the lenses of my own agenda, as a central policy issue: Is fiscal policy used right? And if not, why not?

Why is the question so important? Let me give two examples.

My figure 1 shows the evolution of the ratio of the structural primary balance to potential GDP—for both the euro zone as a whole and for the average of Germany, France, and the Netherlands—from 2005 to 2017. It tells a well-known story. During the acute phase of the crisis, euro zone countries all ran large deficits, going far beyond automatic stabilizers. But, starting in 2010, faced with a rapid increase in the ratio of debt to GDP, and despite the fact that unemployment was still high and growth still weak, the fiscal priority became debt stabilization. By 2013, the structural deficit of the euro zone had been reduced to less than 1 percent of GDP. This was true even in countries where investors did not appear to worry about debt; the figure shows that the average of the adjustments in Germany, France, and the Netherlands, three countries whose 5-year credit default swap rates never exceeded 150 basis points, was as large as for the euro zone as a whole. This shift to fiscal austerity has been widely criticized and blamed

Figure 1. The Structural Primary Balance / Potential GDP (percent)[a]



Source: Eurostat data.

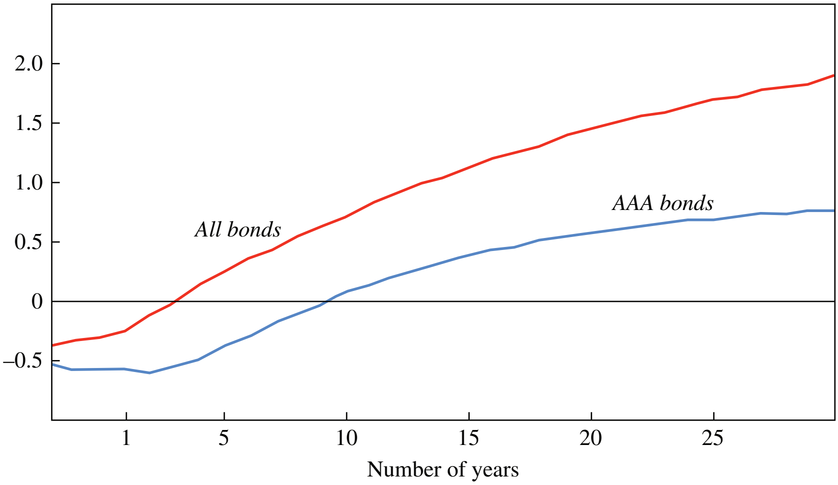
a. This figure shows the evolution of the ratio of the structural primary balance to potential GDP—for both the euro zone as a whole and for the average of Germany, France, and the Netherlands—from 2005 to 2017.

for the slow recovery. The question is why the governments felt that they had to shift priority from output support to fiscal consolidation.

My figure 2 shows the current (April 24, 2019) euro zone yield curves for the average of AAA-rated euro zone government bonds and for the average of all euro zone government bonds.[1] Note how, for AAA-rated bonds, yields are negative up to 9 years, and remain below 1 percent even at 30 years. For all bonds, yields are negative up to 2 years, and remain below 2 percent, even at 30 years. Such low yields are historically unprecedented, and the two curves show the limited space left for monetary policy. If there were to be a slowdown in demand, or an outright recession, it is clear that fiscal policy would need to be used. What are the governments of AAA-rated countries going to do? What are the governments of countries with lower ratings going to do? Will they be willing to rely on fiscal policy as needed, or will they be reluctant to do so, either based on their beliefs or their worries about investors' beliefs and responses? Understanding what they have done in the past, and why they have done it, is clearly of the essence, and it is the focus of Romer and Romer's paper.

1. These two yield curves are constructed by the ECB; see ECB 2019.

Figure 2. Euro Zone Yield Curves[a]



Source: ECB (2019).
a. This figure shows the current (April 24, 2019) euro zone yield curves for the average of AAA-rated government bonds and for the average of all government bonds.

Let me start with the road map of the paper. Romer and Romer first extend their earlier work on the construction of a financial stress index, and on the relation between the index and output, two useful extensions in their own right. They show that (1) higher initial debt is associated with a more conservative fiscal stance, and (2) higher initial debt is associated with a stronger effect of financial stress on output. It is tempting to draw a causal inference: that higher initial debt leads to a more conservative fiscal stance, and thus a stronger effect of financial stress on output. This could be done by regressing output on the financial stress index instrumented by debt. Romer and Romer refrain from doing so (although I am quite sure that they believe that this causal link is relevant), so as to focus on the topic of the paper: Why does debt lead governments to be more fiscally conservative? Is it because they believe that they have no choice, that if they were less so, investors would worry and ask for a large spread, or do they believe they have a choice and conclude in favor of a more conservative fiscal stance? To answer the question, they control in various ways for investors' attitudes; and they conclude, based on both qualitative and quantitative evidence, that more is at play—that debt affects policymakers' decisions, independent of investors' worries.

Is their analysis fully convincing? Before I answer this question, let me get a couple of minor issues out of the way. The first is the issue of how to treat Greece. Greece is an extreme outlier, with respect to level of debt, the size of the fiscal adjustment, and the output shortfall. On one hand, it contains a lot of information; on the other, the circumstances are very specific, and the size of the fiscal adjustment clearly depended not on the Greek government but on the generosity of the European Union and the International Monetary Fund. Romer and Romer recognize the issue in a footnote, and they indicate that some of the results are indeed sensitive to whether Greece is included; it would have been good to report all results both with and without Greece. The second minor issue is that Romer and Romer look only at recessions associated with financial crises. This makes sense, because it builds on their previous work. It could be, however, that those recessions are special. One aspect in particular is the importance of bailouts, which often lead to large fiscal deficits. Bailouts may reflect characteristics of the financial system that affect the recovery, leading to a spurious relation between the budget deficit and the speed of recovery.

Let me turn to what is potentially a bigger issue, whether Romer and Romer convincingly separate what is due to policymakers' worries about investors and what is due to their own beliefs about the dangers of debt. In essence, Romer and Romer run a regression of the form

$$\text{Deficit} = f[\text{debt}(-1), \text{market signal}(-1)], \tag{1}$$

where the two right-hand-side variables are lagged to reduce endogeneity issues. I see two potential issues with the regression. The first is an errors-in-variables in measuring the market signal. The second is the use of the lagged market signal.

The errors-in-variables issue is standard, but likely to be relevant here. Unless Romer and Romer exactly capture the market signal, the coefficient on debt is likely to be biased upward. Let x be the deficit, d be debt, and m be the market signal. Assume the relation reflecting the government decision is given by (ignore the lags here)

$$x = ad + bm + e_x. \tag{2}$$

Assume the market signal depends on debt:

$$m = cd + e_m. \tag{3}$$

And assume we observe a proxy for the market signal:

$$M = m + e_M. \tag{4}$$

If there is no noise, so $e_M = 0$, the regression will indeed have the correct values of a and b, and thus the relative roles of market signals and debt. But the larger the noise, the larger the downward bias on b (this is the standard part) and the larger the upward bias on a, and thus the larger the overestimate of the role of debt. In the limit, if M is fully uninformative, the estimated coefficient on the signal will obviously be equal to zero, and the coefficient on debt will be equal to $a + bc$. Given the difficulty of capturing the correct measure of the market signal, it is likely that Romer and Romer overestimate the direct role of debt.

The other issue has to do with the timing of the market signal in the regression. Take a simplistic example. Assume the relation reflecting the government decision is given by

$$x = bm + e_x. \tag{5}$$

Assume that the market signal depends just on the current deficit, so

$$m = cx. \tag{6}$$

And assume that x is unanticipated, and that, to avoid the simultaneity problem, one runs

$$x = bm(-1) + e_x. \tag{7}$$

Then, in expected value, on average, b will be estimated to be zero, no matter what its true value. In other words, though the policymaker would face a strong market response if he or she were to increase the deficit, the regression would simply miss it. Given Romer and Romer's specification, this may also play a role in finding a limited role for market signals.

Looking forward, I can think of two directions in which to extend Romer and Romer's conclusions. The first is descriptive. It is to look at whether and how rating agencies and investors look at the fiscal position, and how much they worry about debt. The second is normative. If the reluctance of policymakers comes from worries about investors, are the investors right? If it comes from their own worries about debt, are they right? And, if not, what should we, the economists, tell them?

Let me take a pass at each of the two directions. In preparation for writing this comment, I did a literature review of the determinants of ratings of sovereign debt. The variables that appear most often are GDP per capita, past default, the rate of inflation, debt, and deficits.[2] I decided to explore the specification of the determinants of ratings from a study by Goldman Sachs (Ardagna 2018), which considered ratings from Standard & Poor's, Moody's, and Fitch for 21 countries that belonged to the Organization for Economic Cooperation and Development from 1984 to 2017 (subject to data availability for some countries).[3] It mapped each of these ratings into 11 bins, from 1 to 11, with 11 being the best (AAA for all three rating agencies). I then ran an ordered probit for each of the three sets of ratings on a number of variables, the main ones being the log of real GDP per capita, the GDP growth rate, the unemployment rate, the inflation rate, the ratio of the current account balance to GDP, the ratio of the net international investment position to GDP, and two fiscal variables, the ratio of government debt to GDP and the ratio of the primary deficit to GDP. The most consistently significant variables were the log of real GDP per capita and the two fiscal variables, with t statistics above 10 for the ratio of debt to GDP.

To assess the contribution of the ratio of debt to GDP, I replicated the regression and computed—for each country, each year, and each rating agency—the estimated probability that a country had the highest ranking. I then plotted, as shown in my figure 3, the estimated probability against the ratio of debt to GDP. The figure yields a simple and strong conclusion: Relatively low debt is necessary but not sufficient to obtain the highest ranking with high probability:

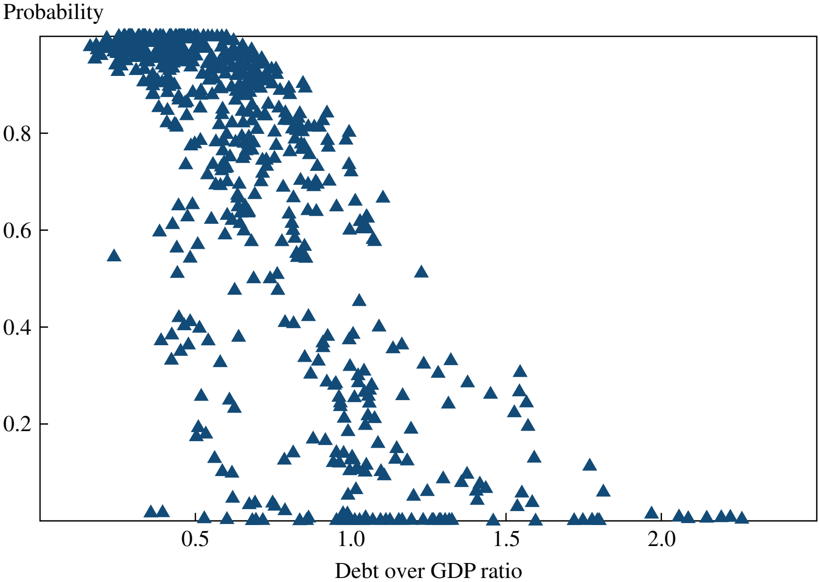
Look first at the outer envelope of the set of points. A necessary condition for obtaining the highest ranking with a probability close or equal to 1 is to have a debt-to-GDP ratio below approximately 70 percent. As debt increases beyond this level, the probability of obtaining the highest ranking decreases to approximately 70 percent for a ratio of debt to GDP of 100 percent, and 30 percent for a ratio of debt to GDP equal to 150 percent.

Look, however, at the set of points below this envelope. It is clear that, for many countries, a low debt-to-GDP ratio is not sufficient to ensure a high rating, and other factors dominate. The conclusion: To the extent that

2. See, for example, Afonso, Gomes, and Rother (2011), which is fairly representative.

3. I thank Silvia Ardagna, the author of the study, and Goldman Sachs, for allowing me to use their data.

Figure 3. The Probability of a Country's Highest Ranking as a Function of the Debt Ratio[a]



Source: Goldman Sachs data.

a. This figure shows the estimated probability that a country had the highest ranking against the ratio of debt to GDP. After I replicated the regression discussed in the text and computed the estimated probability for each country, each year, and each rating agency, I plotted this probability.

ratings get reflected in spreads, it suggests that policymakers are right to think that investors care about debt-to-GDP ratios, and to take this into account.

Turning to the normative implications, are investors and policymakers right to be so worried about debt? Were policymakers right to shift to fiscal austerity in 2011? Should they be willing to sustain potentially large deficits if demand weakens in Europe in the near future? Romer and Romer's paper discusses these issues briefly in the conclusion, but this is a much larger topic, which deserves another paper and another discussion. I made a first pass at it in my 2019 address to the American Economic Association (Blanchard 2019), and, based on that address, you can predict my answers: In an environment where interest rates are low, indeed lower than growth rates, the fiscal costs of debt are small, and so are the welfare costs. And if the effective lower bound prevents monetary policy from decreasing the interest rate as much as required, large deficits can reduce output gaps, and they can have benefits that largely exceed their costs. Fiscal policy was too tight in Europe starting in 2011. Europe should be ready to accept larger fiscal deficits if demand slows down.

The debt numbers on which policymakers, ratings agencies, and investors base their decisions may have been right for an earlier environment, when interest rates were higher and monetary policy was unconstrained. But these debt numbers are no longer the right ones. Today, it is our responsibility to convince all actors to reassess, and to adjust fiscal policy accordingly.

REFERENCES FOR THE BLANCHARD COMMENT

Afonso, António, Pedro Gomes, and Philipp Rother. 2011. “Short- and Long-Run Determinants of Sovereign Debt Credit Ratings.” *International Journal of Finance and Economics* 16, no. 1: 1–15.

Ardagna, Silvia. 2018. “Rating Sovereigns: More Upgrades on the Horizon.” *Goldman Sachs Economics Research Report*, March 16.

Blanchard, Olivier. 2019. “Public Debt and Low Interest Rates.” *American Economic Review* 109, no. 4: 1197–1229.

ECB (European Central Bank). 2019. “Euro Area Yield Curves.” https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates/euro_area_yield_curves/html/index.en.html.

COMMENT BY

PHILLIP SWAGEL I am confident that I was asked to discuss this excellent and insightful paper because of my policy experience, and embrace that in these comments; note, however, that this discussion came before I was appointed as director of the Congressional Budget Office. The key result of Christina Romer and David Romer’s paper is that countries with lower ratios of debt to GDP take more expansionary fiscal action and do better after a financial crisis. They show this using a clever mix of quantitative and qualitative analysis, with data that encompass more countries and years than their past research that shows a similar result.

Romer and Romer thus imply that countries would be wise to ensure that their debt ratios are low enough so that they can use fiscal policy in the face of a downturn. In the parlance of the International Monetary Fund, the recommendation is that countries should maintain “room to maneuver”—that is, to ensure that they have the “fiscal space” to undertake a debt-financed fiscal expansion. Fiscal space is the key concept here, corresponding to the ability to borrow without unnerving markets. A country without fiscal space might not be able to borrow without self-defeating negative effects, such as crowding out from higher interest rates or, in the extreme, risking a debt-driven crisis.

A concern with the empirical work is that the debt ratio might be quite a rough approximation of the measure of fiscal space that the authors really have in mind. Country fixed effects mean that the empirical work accounts for the reality that investors and politicians get nervous about different debt levels across countries. A ratio of debt to GDP that raises concerns for, say, Spain (at least before the European Central Bank's "whatever it takes" bond buying spree), might not pose a challenge for France or Germany, and this cross-country difference might be stable over time. But it could be instead that the threshold at which market participants get nervous about a country's debt burden changes from year to year, in which case the debt ratio would not capture changes in the fiscal ceiling that defines the amount of fiscal headroom. Coming up with a better definition of fiscal space is a worthy goal for future research.

What is both new and of considerable importance for policymakers is the second part of Romer and Romer's empirical work, which concludes that countries with high debt ratios do not take fiscal action apparently because of the constraint of the debt ratio itself. This is the case even controlling for measures of market access such as credit default swap premiums and rating agencies. It is not just that countries facing skeptical investors cannot borrow to fund a countercyclical fiscal expansion, but that they choose not to undertake the fiscal expansion even when they seemingly could. The fiscal constraint, according to Romer and Romer, is a self-imposed mistake: countries with lower debt ratios do better in the wake of a crisis not just because they have the fiscal space but also because they are willing to use it.

Again, the results must be taken with some caution. It could be, for example, that policymakers are cognizant of nonlinear effects that do not come out in the empirical research. Perhaps policymakers recognize that the risks are amplified if they go beyond a certain debt threshold (such as in certain situations not captured by the regressions). In this case, what the paper interprets as caution reflecting ideological blinders might actually be a perceptive insight about the risks of increased debt, perhaps born out of the experiences of emerging markets affected by rapid shifts in investor sentiment. To their credit, Romer and Romer discuss the possible caveats, even while still concluding that countries over the past several decades have hesitated to use fiscal policy and have had worse results not just because of the constraints placed on them by bond markets.

Once a crisis happens, Romer and Romer would instead urge countries to use their available fiscal space to bring about a shallower recession and more rapid recovery. As the authors highlight, the ability to deploy expansionary fiscal policy is especially valuable in the wake of a financial crisis, when monetary policy might be nearing its conventional limits and even quantitative easing might press up against waning effectiveness in boosting investment and consumption.

The target for Romer and Romer's recommendations includes both policymakers deciding on a response in the wake of a financial crisis and officials at multilateral lenders such as the International Monetary Fund, who have considerable influence on whether a debt-financed fiscal expansion takes place in countries that turn to the global lender. The IMF in particular, according to the interpretation in this paper, should be more willing to fund fiscal expansion and only later impose discipline.

I suspect IMF officials would say that they are keenly aware of the benefits of a fiscal expansion but must balance this against the negative effects of running up against the fiscal ceiling. In countries with an especially high debt ratio and thus no fiscal room, I suspect the IMF's line would be that the alternative to the fiscal adjustment of an IMF program is worse rather than better: that interest rates would soar and/or the exchange value of a currency would sink if the IMF financed a fiscal expansion in a country that was at its fiscal ceiling. As Kenneth Rogoff put it in 2002, when he was the IMF's chief economist, "Governments typically come to the IMF for financial assistance when they are having trouble finding buyers for their debt and when the value of their money is falling," and "that when an almost bankrupt government fails to credibly constrain the time profile of its fiscal deficits, things generally get worse instead of better" (Rogoff 2002).

Still, there will be cases that are less extreme than that of an "almost bankrupt" country, and Romer and Romer's policy point would be to nudge the IMF staff to think more about the positive effects of a fiscal expansion in the wake of a financial crisis. This seems like a constructive critique for the IMF staff.

The results of Romer and Romer's paper also connect to the recent policy writings of Olivier Blanchard (2019) in stating that low interest rates imply the possibility of more fiscal space than previously believed, and to those of Jason Furman and Lawrence Summers (2019) in suggesting that the United States should use some of its fiscal capacity for productive purposes (they point to government spending programs that might be seen as investments relating to children, health care, and education, though in principle their logic applies to debt-financed tax cuts). Romer and Romer's results suggest, to the contrary, that countries should focus more on fiscal adjustment so they have the fiscal space to use when needed—to fix the roof while the sun is shining. But the key again is to gauge the available fiscal space. A country with a seemingly high debt ratio might have the fiscal headroom to undertake new spending while maintaining the ability for a postcrisis expansion. This points again to the value of further research on gauging countries' fiscal space.

REFERENCE FOR THE SWAGEL COMMENT

Blanchard, Olivier. 2019. "Public Debt and Low Interest Rates." NBER Working Paper 25621. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25621.

Furman, Jason, and Lawrence H. Summers. 2019. "Who's Afraid of Budget Deficits? How Washington Should End Its Debt Obsession." *Foreign Affairs*, March–April. https://www.foreignaffairs.com/articles/2019-01-27/whos-afraid-budget-deficits.

Rogoff, Kenneth. 2002. "An Open Letter to Joseph Stiglitz." International Monetary Fund. https://www.imf.org/en/News/Articles/2015/09/28/04/54/vc070202.

GENERAL DISCUSSION Steven Davis wondered to what extent policymakers' choices were driven by concerns about future loss of market access, and whether these concerns were justified in influencing policymakers' decisions. He asks the authors to characterize more fully the ideas that drove policymakers to not utilize fiscal space more fully.

Valerie Ramey said that she supports the authors' decision to not pursue some of the discussants' recommendations. She complimented the paper for contributing comprehensive, descriptive information. However, she thought that drawing any causal relationships may be problematic because there is no identification in the paper. For instance, she refers to the authors' note that excluding the contemporaneous relationship between GDP and financial distress reduces the negative aftermath of a crisis by nearly half. She observed that this suggests that endogeneity issues are indeed important, so is difficult to ascribe the greater fall in output from countries with a high debt-to-GDP ratio to the government spending response.

Ramey then addressed the recommendation that the authors use debt as an instrument to examine whether fiscal policy affects the speed of recovery. She thought that the authors were correct in not pursuing this because she did not think that they had all the variables to do this with rigor. She made an analogy: say that doctors treat only one of two patients with the flu, and that the treated patient gets very sick, while the other patient does not. If one does not have all the appropriate measurements, it is easy to assume that when doctors are more proactive, the patient gets sicker. However, the different patient outcomes could be due to other underlying conditions. For instance, the treated patient may have been much older. In the authors' case, Ramey did not think that they had all the controls for a causal analysis, even after including country fixed effects, because the fixed effects would not account for everything.

She also supported the discussants' comments regarding the importance of forward-looking behavior. She did not think that current credit ratings entirely capture the effect when the debt-to-GDP ratio increases. She referred to Olivier Blanchard's figure 3, which plots the probability of having the highest debt rating as a function of the country's debt ratio, and she said that it shows the trend that people worry about: there is a steep decline in the probability of having the highest ranking when the debt-to-GDP ratio gets too high.

Ramey then mentioned evidence that multipliers in government spending differ according to the debt-to-GDP ratio. According to a paper by Ethan Ilzetski, Enrique Mendoza, and Carlos Végh, multipliers range from close to zero to negative if the debt-to-GDP ratio is above 0.6.[1] In these cases, she remarked, an increase in government spending would actually be counterproductive.

Frederic Mishkin said that the key to recoveries from financial crises is recapitalizing the financial system. The issue of fiscal space is important in this regard, he said, because if a country is worried that it does not have enough resources to recapitalize the system, it will not do so. He pointed to the stress tests in the United States versus the stress tests in Europe, and claimed that the tests were one of the most important elements that helped the United States recover from the global financial crisis faster than other countries.

For stress tests to be effective, he explained, there should be a backstop with fiscal policy to provide the funds. Otherwise, if there are poor test results without a backstop, it will exacerbate the crisis. Further, he said, the stress tests have to be credible: an important explanation for the different outcomes in Europe versus the U.S. was the credibility of the stress tests. In the U.S., stress tests were trusted and banks were recapitalized, and that

1. Ethan Ilzetzki, Enrique G. Mendoza, and Carlos A. Végh, "How Big (Small?) Are Fiscal Multipliers?" *Journal of Monetary Economics* 60, no. 2 (2013): 239–54.

is when the recovery started to take hold. Stress tests in Europe, conversely, were very poor. For instance, European stress tests indicated that Anglo-Irish Bank was sufficiently capitalized. Subsequently, Anglo-Irish had to be bailed out at a huge cost to taxpayers. The Europeans were afraid that if they showed that an bank like Anglo-Irish was in trouble, then the crisis would get much worse.

He ended by saying that policymakers should want fiscal space, not necessarily so that governments can afford things like the rail system in California, but to provide the ability to recapitalize the financial system very quickly.

Wendy Edelberg observed that policymakers must weigh the trade-off of the countercyclical effects of a stimulus versus the increasing risk of a financial crisis from high and rising debt-to-GDP. She wondered whether the authors had a hypothesis as to which cost component policymakers were weighting incorrectly: are they putting too little weight on the countercyclical effects or too much weight on the increasing risk of a financial crisis? Or, she wondered, is it possible that they are putting too much weight on some positive aspect, such as a countercyclical effect of improved access to markets?

Donald Kohn observed that the themes arising from previous comments were about fear of losing market access or about rising cost. He remarked that the two fears are not completely separate, and that they are similar in that they fear the same thing.

He then wondered what the authors thought about the role of current account deficits. Particularly, he wondered whether the dependence on foreign inflows, and the concern that foreign investors may be more sensitive than domestic investors to deteriorating debt-to-income ratios, play a role in the way that governments and politicians respond. He asked if governments and politicians worried about a sudden stop in the inflows.

Natasha Sarin wondered about the underlying motivation behind policy-makers' choices in responding to the crises. She wondered if it was mostly political, or if they understood what the implications would be for investors in the market.

Jay Shambaugh wondered how the authors would recommend warning countries when they are more likely to wind up in stress. For instance, he said, a prescription could be to tell governments to use their space when it is there, and to tell countries that are known to be safe that they are.

He then wondered whether the currency of issue played a role in the crisis response. Although the authors said that the debt-to-GDP ratio was related to an increase in sovereign stress, he noted that their scatterplot figure shows that this is only true for countries in the euro area. The pattern was not true for the United Kindom, Japan, and the United States. Shambaugh commented that this may be because these countries were issuing in a currency that they could print, as opposed to countries in the euro area.

Jason Furman claimed that he would not be defending the Obama administration, nor the Californian high-speed railroad, which he said was a decision foisted upon some policymakers. He then humorously said that he would take advantage of Larry Summers's absence to reinterpret Furman and Summers's paper in a way that he himself would have meant it.

In their paper, he claimed that they were very clear that fiscal policy undertakings should be paid for. The debate is about *what* should be paid for. In his opinion, the Congressional Budget Office (CBO) does a very good job in providing judgment for this decision. He did not think that the CBO would make the mistake of claiming that high-speed rail would have miraculous growth and pay for itself. But the CBO would have accounted for the macroeconomic impact of the fiscal stimulus in alternative analyses that supplement its formal estimates. In fact, he noted that the CBO did account for the minimal growth effects of the 2017 tax cuts. Furman thought that we should have a standard like the CBO's rather than using our own subjective judgments when making these decisions.

He then asked the authors how much of their results were about absolute values versus relatives. He explained that for any given country, the distinction may not matter. But for the International Monetary Fund's advice to the world as a whole, it would matter. For instance, a country may be treated as responsible if its debt ratio is not much higher than other countries' debt ratios. A country with a ratio of 140 percent on its own is in a different situation than if the world becomes accustomed to 140 ratios as a whole.

Martin Baily believed that one of the main reasons why there was not more expansionary fiscal policy is that many decisionmakers did not believe that it would be effective. Many German policymakers thought Keynes was wrong and that fiscal multipliers were not positive. He said the U.S. economics profession, which is not united on this issue, did not persuade some countries that Keynesian fiscal policy would be an effective response to the crisis. The Harvard economist Alberto Alesina (who has said that spending cuts to reduce budget deficits have frequently been followed by economic growth), was quoted often in Europe.[2] And the

2. Alberto Alesina and Silvia Ardagna, "Large Changes in Fiscal Policy: Taxes versus Spending," *Tax Policy and the Economy* 4, no. 1 (2010): 35–68.

European Commission also had studies indicating that fiscal multipliers were negative and held the belief that reducing budget deficits would be the best way to return to full employment.

He explained that one reason for this is that the economics profession, which is not united on this issue, did not persuade some countries that expansionary fiscal policy would be an effective response. In Washington during the crisis, Baily would ask people in closed meetings whether they thought Keynes was dead, or if they thought they just could not implement the policy. Respondents either refused to answer, or thought that Keynes was dead—they did not believe that it would work. Further, the European Commission had studies indicating that fiscal multipliers were negative, and held the belief that fiscal contraction would be the best way to return to full employment. His final point: it is important to keep in mind that people just did not understand, or did not believe, that the fiscal multipliers were positive. They thought they were negative.

Stanley Fischer noted that when studying the dynamics of what happens after crises, an important variable is whether a country had a financial crisis or not. In fact, the authors have written about this topic, he said. He wondered if the authors attempted to explain the need for fiscal space in terms of whether a country had a financial crisis, and whether the fact that there was a financial crisis in a country was an important determinant of the subsequent dynamics.

Gerald Cohen noted that the IMF gave a sort of backstop facility to countries like South Korea early in the global financial crisis. He wondered if the IMF programs or the Federal Reserve's foreign exchange swap lines were also a sort of backup access to dollar-denominated debt. He asked if this information could be used to see if countries had extra fiscal space that was otherwise not indicated by other measures.

Robert Hall noted that they had been discussing debt in the traditional sense by examining the debt-to-GDP ratio. But another way to view debt is to look at the ratio of debt service to GDP, he said, and he expected that this measure would produce a very different answer. For instance, in Germany, debt is actually a revenue source with negative yield. And in the United States, there has been an explosion of debt, but debt service has been quite low. This generates a big risk if interest rates were to rise, he said. But according to the paper by Rachel and Summers, that is not where interest rates seem to be headed.[3] He said that he would be interested in knowing how the authors' findings would change if they examined the debt service-to-GDP ratio.

3. See the paper by Rachel and Summers in this volume of *Brookings Papers*.

Jeromin Zettelmeyer noted that it is inherently difficult to distinguish between a market access story and the voluntary choice story, because part of what may be driving these voluntary choices by policymakers is the concern that borrowing conditions may deteriorate in a crisis situation with fiscal expansion. That is, in the extreme, what losing market access means, he said. He observed that this certainly motivated some decisions made at that time by European countries, which did not pursue fiscal expansion on the grounds that it was thought to have a negative effect.

David Romer began by thanking the commenters Olivier Blanchard and Phillip Swagel for their presentations and those who contributed to the discussion. Janice Eberly made an aside that the authors had anticipated all the discussed concerns in the footnotes.

Romer first addressed the comments by Blanchard, who had wondered if the authors could, in more detail than in their footnote, explain what happens to their results when Greece is excluded from the sample. Romer acknowledged that the footnote did not do a sufficient job in reporting those exercises. As expected, he said, including Greece strengthened their results, but it is clear that Greece does not drive the conclusions.

He agreed with Blanchard's concern regarding measurement error in the market access variable. The authors tried hard to address this, he said, and considered multiple different measures in their analysis. In addition, he observed that the debt ratio is a noisy measure of what is driving policymakers' ideas about fiscal space, which would tend to bias the results in the opposite direction from measurement error in the measures of market access. Thus, it is not clear which way measurement errors on net bias the results. Indeed, Romer noted that one of the motivations for doing the narrative work in the paper was that it is another way of getting access to this information: intelligent observers may have a better sense of what true market access was. The narrative analysis suggests a larger role for market access than the statistical work does, but it still shows a large role for policymakers' ideas.

He referred to Blanchard's comment that lagged market access is a noisy measure of current market access, and thus, to the extent that market access is not serially correlated, the use of lagged market access potentially biases the coefficient on market access. The measures of market access are in fact substantially serially correlated, Romer said, but he acknowledged the concern.

Romer then addressed a broader set of issues touched upon by both discussants and commenters. Many had asked about the specific underlying motivations driving policymakers' choices. In the paper's narrative source, Romer said, many justifications for fiscal tightening referred to a potential loss of market access at some vaguely distant point in the future (he clarified, in response to Zettelmeyer's observation, that statements like these were categorized as reflecting market constraints rather than policy choices). However, there was no evidence in any of their measures, nor from the Economist Intelligence Unit, that countries were actually in danger of losing market access in these cases. As to whether policymakers have a deep insight into economic conditions and subtle threats to market access, Romer said he did not know. But he pointed to Martin Baily's observation about decisionmakers not believing that fiscal stimulus would be an effective policy response, and Romer said that this was a different potential source of policymakers' choices and a valuable channel to explore.

Romer then referred to Valerie Ramey's impassioned speech about being careful about ascribing causal relationships. He agreed with her comments about not drawing strong causal conclusions from the correlations between financial distress and the behavior of GDP, and about the pitfalls of pursuing an instrumental-variables approach to the impact of the debt-to-GDP ratio. But he also pointed out that she cited the paper by Ilzetski, Mendoza, and Végh, which, to his recollection, lacked careful treatment of identification. The paper focused on reduced-form regressions, and Romer argued that Ramey therefore could have given the same speech about that paper.

Christina Romer contributed to the inquiries wondering what the specific ideas driving policymakers' decisions were. She seconded Martin Baily's idea that many decisionmakers after the 2008 crisis just did not believe that fiscal stimulus would work. Indeed, she remembered that Alesina and Ardagna's ideas about fiscal policy carried a lot of weight at the time.

She also mentioned that although she believed the stress tests were helpful in 2009, she thought their contribution might be overblown. She worried that a narrative was developing that the differences among countries' postcrisis outcomes are not because of their different fiscal expansions, but because some countries had credible stress tests while others did not. This perpetuates the view that fiscal policy did not matter, she said. She again highlighted the finding of the paper that countries with more fiscal space do more fiscal expansion for both financial rescue and countercyclical actions, and have significantly better outcomes. She emphasized that this relationship should not get lost in the discussion.

David Romer noted that another channel through which fiscal space may work is that it may prevent a country from having financial distress in the first place. He pointed to the mechanism that Frederic Mishkin described: if the financial system starts to come under stress, it will not break down if people know that an entity would step in and bail the system out so that it could continue to function smoothly. However, he wondered if it was possible to gather formal evidence on this. Nonetheless, he said, the idea suggests that at the very least, there is one channel through which their results were understating the benefits of having fiscal space—perhaps having fiscal space allows countries to not need to use it.

STEPHANIE R. AARONSON
Brookings Institution

WILLIAM L. WASCHER
Federal Reserve Board

MARY C. DALY
Federal Reserve Bank of San Francisco

DAVID W. WILCOX
Federal Reserve Board (Retired)

Okun Revisited: Who Benefits Most from a Strong Economy?

ABSTRACT Previous research has shown that the labor market experiences of less advantaged groups are more cyclically sensitive than the labor market experiences of more advantaged groups; in other words, less advantaged groups experience a high-beta version of the aggregate fluctuations in the labor market. For example, when the unemployment rate of whites increases by 1 percentage point, the unemployment rates of African Americans and Hispanics rise by well more than 1 percentage point, on average. This behavior is observed across other labor market indicators, and is roughly reversed when the unemployment rate declines. We update this work to include the post–Great Recession period and extend the analysis to consider whether these high-beta relationships change when the labor market is especially tight. We find suggestive evidence that when the labor market is already strong, a further increment of strengthening provides a modest *extra* benefit to some disadvantaged groups, relative to earlier in the labor market cycle. In addition, we provide preliminary evidence suggesting that these gains are somewhat persistent for African Americans and women.

Conflict of Interest Disclosure: Stephanie Aaronson is the vice president and director of the Economic Studies program at the Brookings Institution; Mary Daly is the president and chief executive officer of the Federal Reserve Bank of San Francisco; William Wascher is deputy director of the Division of Research and Statistics of the Federal Reserve Board of Governors; and David Wilcox is the former director of the Division of Research and Statistics of the Federal Reserve Board of Governors. 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 Brookings Institution, the Federal Reserve Bank of San Francisco, or the Federal Reserve Board of Governors.

The difference between unemployment rates of 5 percent and 4 percent extends far beyond the creation of jobs for 1 percent of the labor force.

—Arthur Okun, *Brookings Papers on Economic Activity*, 1973

In 1973, Arthur Okun wrote an iconic paper asking whether a "high-pressure economy" could contribute to the upward mobility of U.S. workers. Okun's hypothesis was simple. In a high-pressure economy—defined by resource utilization running beyond its longer-run sustainable rate—firms would find it difficult to fill vacancies at a given wage and would react by relaxing hiring standards and reducing their use of statistical metrics for evaluating candidates in favor of more intense personal screening.[1] He argued that these changes had the potential to improve the economic circumstances of less advantaged workers, allowing them to find employment, build their skills, and climb the job-and-income ladder. Looking at the data, he found that these benefits were indeed a feature of high-pressure periods in U.S. economic history; during high-pressure episodes, men moved up the job ladder, creating room for women and teenagers to move into the labor market. On the basis of these findings, Okun concluded that though not a panacea, a high-pressure economy complemented other policies working to achieve the social objective of upward mobility.

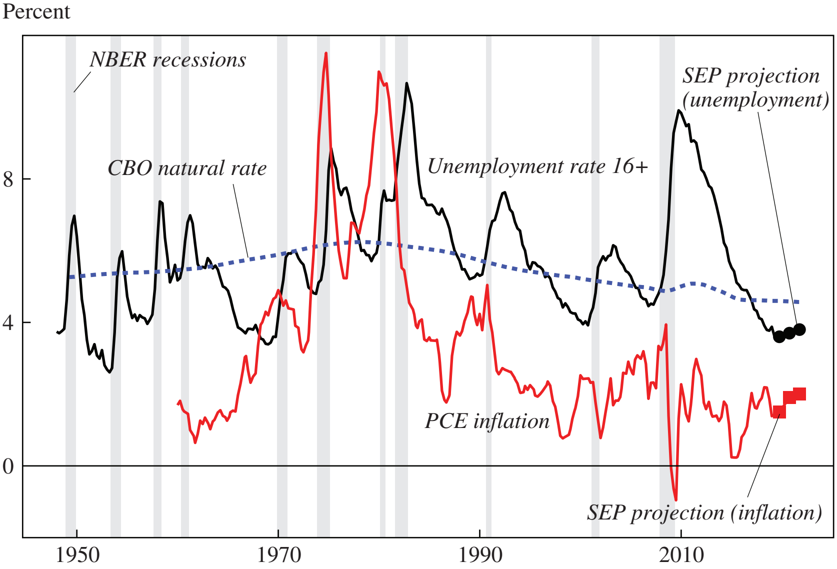
Nearly 50 years later, Okun's analysis remains relevant.[2] The current economic expansion has now become the longest in U.S. history and the labor market is tight by most standards. Moreover, inflation has been muted, running consistently below the 2 percent target of the Federal Open Market Committee (FOMC). As shown by the heavy solid line in figure 1, the unemployment rate, a standard measure of labor market strength, is currently about as low as it has been since 1969. Moreover, it is well below the estimate by the Congressional Budget Office (CBO) of its longer-run sustainable value (the dotted line).[3]

1. See Okun (1973, 240).

2. In the fall of 2016, the minutes of FOMC meetings and then–Federal Reserve chair Janet Yellen noted the emerging debate about the potential of running a "high-pressure economy." This discussion has continued in the media and publicly since that time and has been among the topics at the series of Fed Listens events held in 2019; see Federal Reserve Board of Governors (2019b).

3. The CBO's views are aligned with those of private sector forecasters (as measured by the Blue Chip consensus) and the FOMC's "Summary of Economic Projections" (SEP); as of March 2019, the CBO's estimate of the natural rate of unemployment was about 4½ percent, while the medians from private forecasters (Blue Chip) and the SEP were at 4¼ percent—all quite a bit higher than the actual unemployment rates that have prevailed

Figure 1. Unemployment and Inflation, 1950–2021[a]



Sources: Bureau of Labor Statistics; CBO; SEP, March 2019.
a. CBO = Congressional Budget Office; NBER = National Bureau of Economic Research; PCE = Personal Consumption Expenditures Price Index; SEP = Federal Open Market Committee's "Summary of Economic Projections."

Looking ahead, based on the median of the FOMC's March 2019 "Summary of Economic Projections," indicated by the dot symbols on the heavy solid line in figure 1, the unemployment rate is expected to remain below 4 percent through 2021.[4] If this forecast is borne out, the U.S. unemployment rate will spend much of the next few years ½ percentage point or more below the CBO's estimate of its long-run sustainable level. Although the unemployment rate does move below the CBO's estimate of its sustainable level (a negative unemployment gap) with some regularity, a high-pressure expansion of this duration would border on exceptional.

The experiences of a high-pressure economy at various points over the past 40 years afford an opportunity to revisit Okun's question and to

over the past year. The labor market strength seen by economists and policymakers is also reflected in surveys of households and firms. In the Conference Board's Consumer Confidence Survey, for example, a much larger percentage of respondents stated that jobs are plentiful than said that jobs are hard to get, while in the National Federation of Independent Business's survey of small businesses, the percentage of companies reporting that jobs are hard to fill is at a historically high level.

4. See FOMC (2019).

document who benefits most from a strong economy. In particular, we are interested in the degree to which less advantaged groups of workers see disproportionate improvements in employment and income when the labor market is especially tight. We add to the existing literature by updating the analysis to include the current expansion, to focus specifically on whether the dynamics of key variables differ during hot labor markets, and to consider both the short- and longer-term impact of high-pressure periods on less advantaged groups. We also consider whether rural areas do better or worse than urban areas and whether the results hold in metropolitan-area-level, rather than national, data.

The analysis demonstrates several important points. We reaffirm the earlier findings of other authors that the labor market outcomes of blacks, Hispanics, and those with less education are more cyclically sensitive than the outcomes of whites and those with more education. We find that this greater cyclical sensitivity holds in both cold periods (those with a positive unemployment gap) and hot periods (those with a negative unemployment gap). Moreover, we find suggestive evidence that when the labor market is already strong, certain groups of disadvantaged workers benefit even more than usual from further strengthening. In other words, for these groups the last increments of strengthening appear to reduce labor market disparities by even more than earlier increments of strengthening had done. Notably, for prime age workers, these gains appear to be at least somewhat persistent along the participation rate dimension.[5]

The bulk of our inquiry focuses on individuals age 25 to 64 years; however, we also briefly examine data for younger persons, age 16 to 24, and find that the labor market experiences of young black workers are more cyclically sensitive than are the experiences of white youths and blacks age 25 to 64.

In contrast to the results for unemployment and participation, we find little evidence that gaps in hourly wages, annual own earnings, and household income vary over the labor market cycle; when they do change, they tend to widen. These results are consistent with previous research by Hilary Hoynes (2000); Jonathan Parker and Annette Vissing-Jorgensen (2010); Mary Daly, Bart Hobijn, and Joseph Pedtke (2019); and Cynthia Doniger (2019).

5. Reifschneider, Wascher, and Wilcox (2015) show that the presence of hysteresis is a relevant consideration for monetary policymakers.

The remainder of the paper is organized as follows. Section I provides a summary of the existing literature. Section II describes the data and measurement of key variables. Section III reviews the results on the relative sensitivities of important groups across key labor market and income indicators—including unemployment rates, labor force participation rates, wages, and household incomes. Section IV discusses some potential costs of running a high-pressure economy that policymakers should consider, and section V offers tentative conclusions from our investigations.

I. The Previous Literature

Following Okun (1973), many authors have investigated elements of the high-pressure hypothesis. A number of studies written in the wake of the strong economy of the late 1990s documented that disadvantaged workers, including blacks and low-skilled workers, experienced greater cyclical variation in their labor market outcomes. One example is the paper by Hoynes (2000), who examines how the employment, earnings, and income of less-skilled men vary over the business cycle. She finds that men with lower levels of education and nonwhites experience greater cyclical fluctuations in employment and earnings than high-skilled white men, but that earnings of other family members and government transfers mute the impact on family income.[6] Another prominent example is Lawrence Katz and Alan Krueger's (1999) exploration of whether the distributions of wages and incomes tighten systematically as the economy strengthens. They find that the wage growth of lower-wage individuals is more responsive to reductions in the unemployment rate than is the wage growth of higher-wage individuals, and that the tight labor market of the late 1990s produced more widespread benefits for the disadvantaged than did the tight market of the late 1980s, though this partly resulted from the expansion of the Earned Income Tax Credit during the later period.[7] Christina Romer and David Romer (1999) confirm that U.S. poverty rates decline during economic expansions, but they argue, based on cross-country data, that these are merely short-term benefits and that efforts by monetary policymakers to keep the unemployment rate low at the expense of higher inflation are

6. See also her literature review for a discussion of prior studies focusing on the relative labor market outcomes of workers by race and education.

7. Katz and Krueger also caution that the wage and income gains among low-wage workers and low-income families were not sufficient to overcome the trend increase in inequality over the preceding decade.

detrimental to the long-run well-being of the poor. More recently, Philip Jefferson (2008) has examined the behavior of employment-to-population ratios over the business cycle by level of educational attainment. He finds that the cyclical sensitivity of employment was greater from 1968 to 2005 for individuals with lower levels of educational attainment. Similarly, Tomaz Cajner and others (2017) find that both unemployment rates and patterns of labor force entry and exit for blacks and Hispanics are more cyclically sensitive than for whites.

Fewer studies have focused on the question we address here of whether the dynamics of key labor market variables differ when the economy is hot. One exception is Katherine Bradbury (2000), who, using data from the 1970s through 1990s, finds that the difference between black and white men's unemployment rates is about ½ percentage point smaller in periods when the unemployment rate falls below 5 percent, even after controlling for the state of the business cycle using the GDP gap. She does not find a similar, separate effect on the unemployment rate gap between black and white women. Valerie Wilson (2015) compares the 1990s with several less-robust expansions and shows that with respect to both unemployment and earnings, African Americans particularly benefited from the high-pressure economy of the late 1990s. Julie Hotchkiss and Robert Moore (2018) analyze panel data from the National Longitudinal Surveys of Youth and find evidence that high-pressure economies lead to lower rates of unemployment and higher labor force attachment among disadvantaged groups, but that the effects are not particularly long-lived. Similarly, simulations conducted by Bruce Fallick and Pawel Krolikowski (2018) indicate that a hot labor market has modest but short-lived benefits for the labor market outcomes of less educated men.

In trying to understand these various findings, it is helpful to think about the specific channels through which a high-pressure economy could lead to improved labor market outcomes for more marginalized workers. As conceived by Okun in his seminal work, employers may upgrade workers into more productive jobs during a high-pressure economy, with the result that more marginal workers (women and teenagers, in Okun's analysis) increase their employment. A number of studies provide evidence of this phenomenon. Harry Holzer and others (2006) find that during the tight labor market of the 1990s, employers were more likely to hire workers with some stigma, including welfare recipients and those with little experience, although they were not more likely to hire those with a criminal record. Employers also demanded fewer general skills. This latter finding is confirmed by Alicia Sasser Modestino and others (2016), who, using job-posting data, find that in the immediate aftermath of the Great Recession, employers increased skill requirements listed in job postings, such as education and prior experience, and reduced them as the expansion gathered strength. Paul Devereux (2002) provides evidence that new hires tend to have lower educational attainment when the unemployment rate is low and that low-skilled workers experience the greatest occupational improvement in tight labor markets. This result is consistent with the model of vacancy chains developed by George Akerlof and others (1988), whereby as the unemployment rate falls, workers move into jobs that provide better matches. These studies all suggest that the benefits of a high-pressure economy are greater than those that would result simply from the fall in the unemployment rate.

II. Data and Measurement

Most of the data we use come from the Current Population Survey (CPS)—the survey of households used by the Bureau of Labor Statistics (BLS) to construct estimates of labor market outcomes. We focus our attention on 25- to 64-year-olds because this age group consists of individuals who are most likely to be finished with schooling and below normal retirement age. Within this group, we examine the relative outcomes of historically less advantaged groups defined by race, gender, and educational attainment. We define three mutually exclusive groups for race and ethnicity: African Americans or blacks (we use the terms interchangeably); Hispanics or Latinos (again, we use the terms interchangeably); and non-Hispanic whites. We do not show results for Asian Americans, Native Americans, and others separately due to the statistical unreliability of results for smaller sample sizes. We define three levels of educational attainment: a high school degree or less; some college (which includes individuals with post–high school education who did not graduate from a four-year college, including those who earned an associate degree); and a four-year college degree or more. For annual household income, we take the demographic characteristics of the reference person or "householder" for each household in the Annual Social and Economic Supplements of the CPS.[8] All earnings and income series are deflated by the headline Personal Consumption Expenditures Price Index.[9]

8. We exclude "group quarters" households where the householder is not identified.

9. In all our statistical investigations, we use *gaps* in income between two different groups, constructed as 100 times the difference in log incomes. The choice of price index does not affect these gaps, but it does affect the levels shown in figures 4 and 5.

We also do some robustness checks using data at the metropolitan statistical area (MSA) level. For this MSA analysis, we use the outgoing rotation group files of the CPS beginning in 2004, when the U.S. Census switched to designating geographic areas using the core-based statistical area (CBSA) classification system, and ending in 2018. To ensure that we get a sufficient sample to calculate group-specific labor force status by CBSA, we pool the data to the annual frequency, include men and women together, and include areas with at least 500,000 individuals and at least 75 observations for the particular race/ethnicity/education group being analyzed.

Finally, we define cold and hot periods as those when the aggregate unemployment rate is respectively above or below the natural rate as estimated by the CBO—in other words, when the unemployment rate gap is positive or negative. For the MSA analysis, we define the natural rate in each metropolitan area as the average unemployment rate in the period from 2004 to 2008.

III. Results

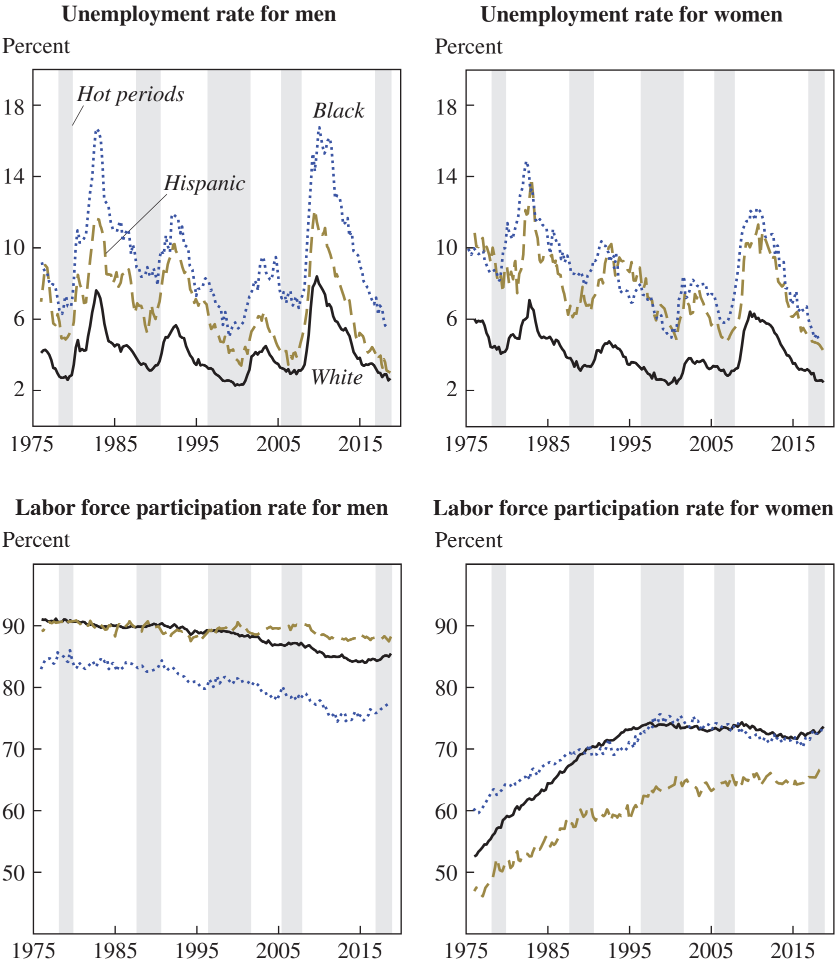
Among the myriad possible labor market outcomes, we focus on five measures: the unemployment rate; the labor force participation rate (LFPR); average hourly wages (which include the wages and salaries of employees, but not the self-employed); annual own earnings (including income from self-employment); and annual household income (from all sources).[10] We compare outcomes for black and Hispanic men and women with outcomes for white men and women; similarly, we compare outcomes for men and women with a high school degree or less and some college to outcomes for men and women with a college degree or more.

III.A. Evidence on the "High-Beta" Experience of Disadvantaged Groups

To set the stage for the results, it is useful to describe the trends in each of the key outcome variables. Figures 2 through 5 plot, in a time-series format, each of the outcome variables for each of our key groups. The gray bars denote periods when the unemployment rate was below the natural rate as estimated by the CBO.

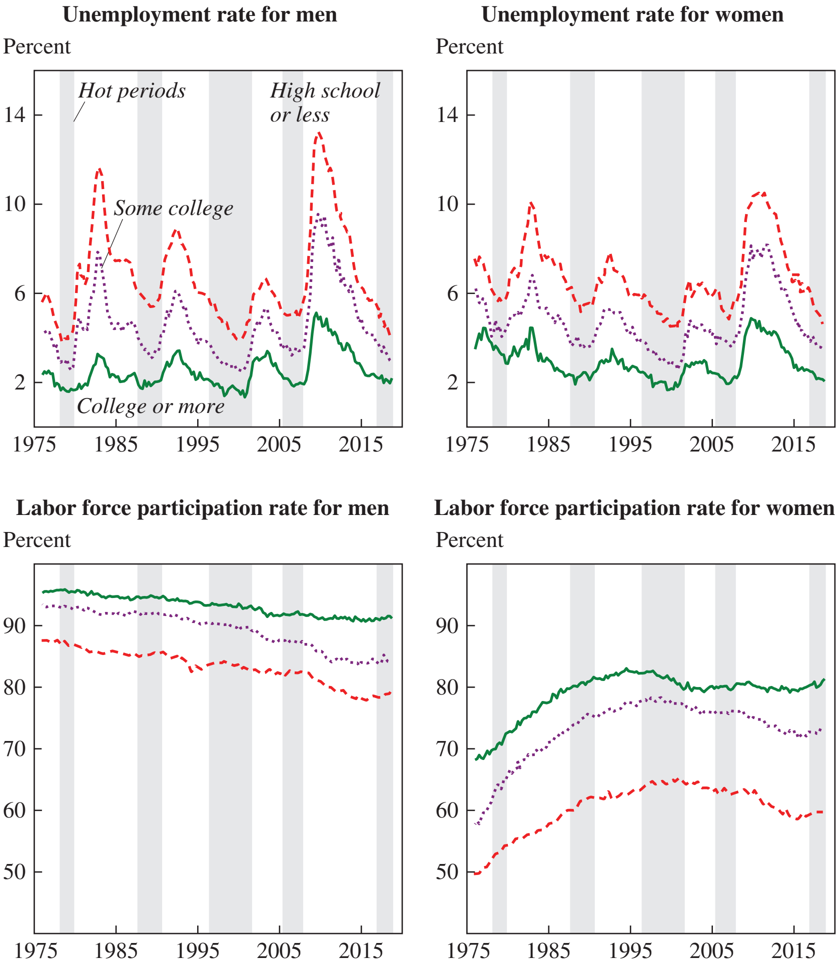
10. For completeness, we perform a similar analysis for the employment-to-population ratio. These results are available in the online appendix. The online appendixes for this and all other papers in this volume may be found at the Brookings Papers web page, ww.brookings.edu/bpea, under "Past BPEA Editions."

Figure 2. Labor Force Statistics by Race and Ethnicity, Age 25–64 Years, 1975–2018



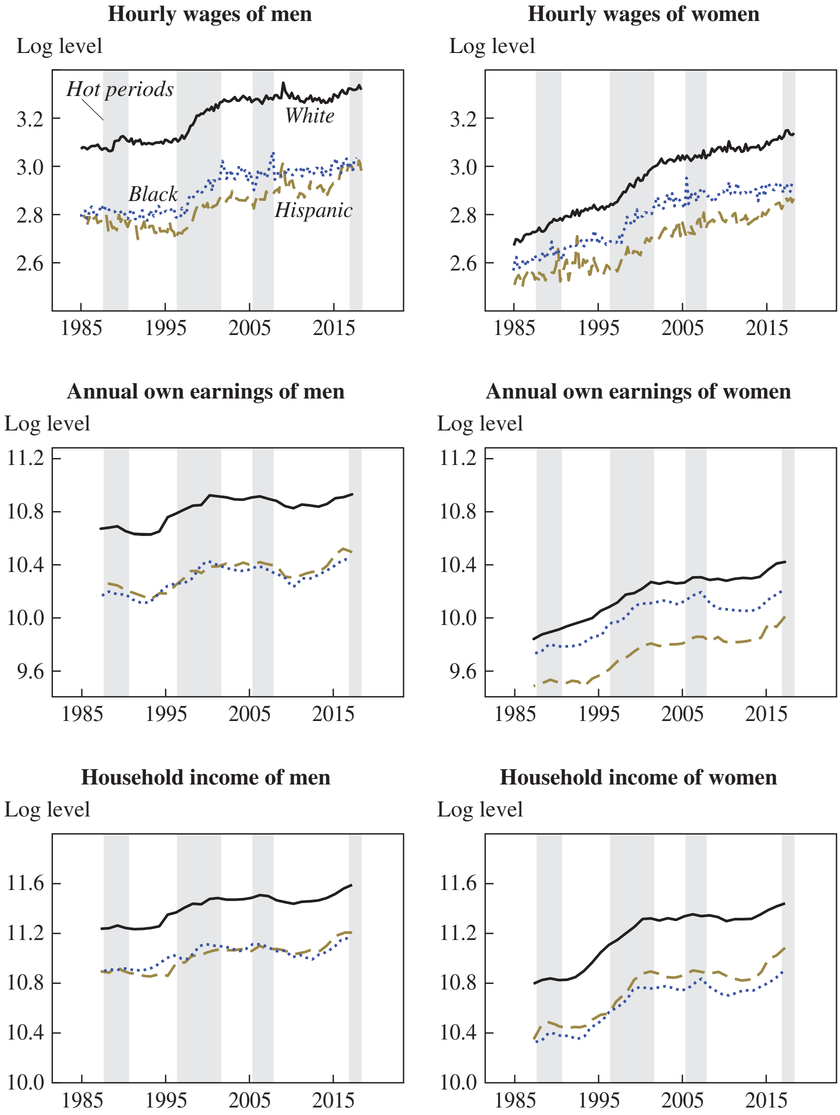
Sources: U.S. Census Bureau; Bureau of Labor Statistics (Current Population Survey).

Figure 3. Labor Force Statistics by Education Level, Age 25–64 Years, 1975–2018



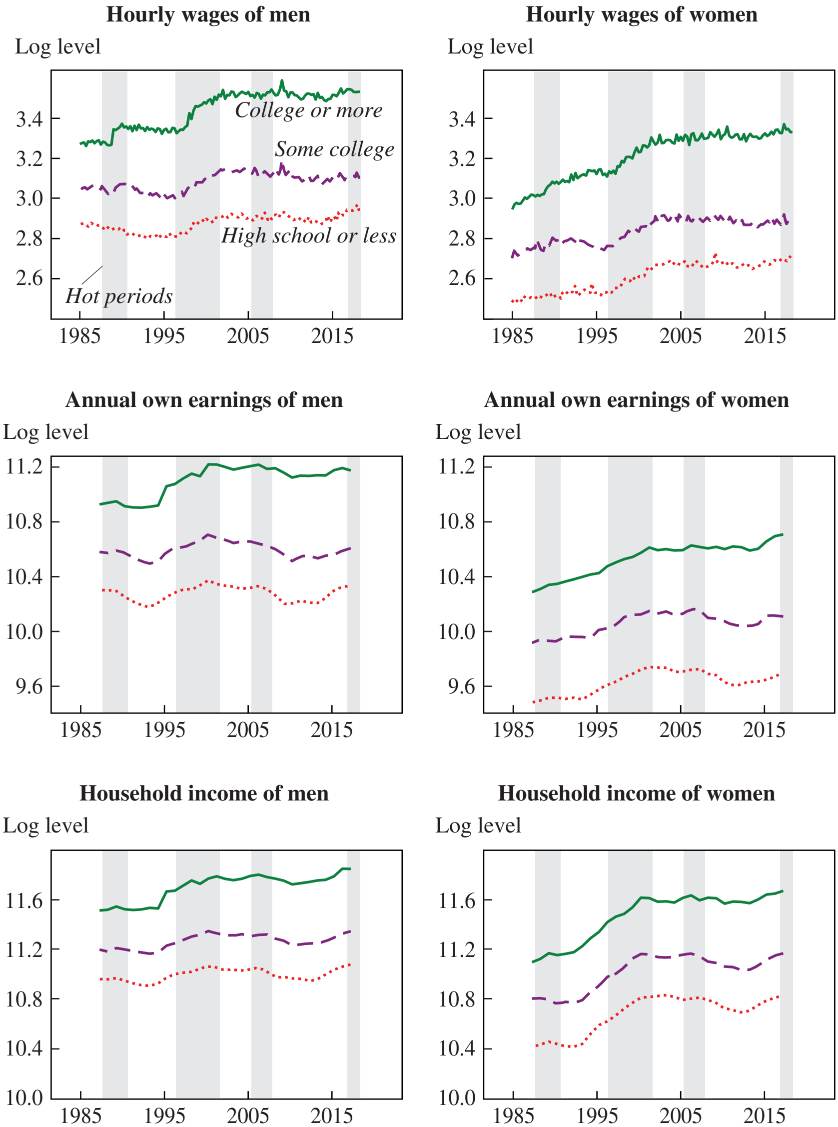
Sources: U.S. Census Bureau; Bureau of Labor Statistics (Current Population Survey).

Figure 4. Earnings and Income by Race and Ethnicity, Age 25–64 Years, 1985–2018



Sources: U.S. Census Bureau; Bureau of Labor Statistics (Current Population Survey).

Figure 5. Earnings and Income by Education, Age 25–64 Years, 1985–2018



Sources: U.S. Census Bureau; Bureau of Labor Statistics (Current Population Survey).

A key feature evident in figure 2 is that fluctuations in the unemployment rates for African Americans and Hispanics—both men and women—are roughly synchronized with fluctuations in the unemployment rate for whites (the top two panels). However, the rates for African American and Hispanic men and women are uniformly higher than the rates for white men and women, and they exhibit considerably greater amplitude. As a result, when the labor market weakens, the gaps between these rates widen markedly; they then shrink again when the labor market tightens.

Compared with the unemployment rate, the LFPR (the bottom panels) is considerably less cyclically sensitive. A much greater fraction of the variation in the gaps in the LFPR across different races and ethnicities appears to reflect secular trends. Overall, black men have a lower LFPR than do white or Hispanic men. Among women, Hispanics participate at a lower rate than do either blacks or whites.

Figure 3 presents similar information for groups at different levels of educational attainment. On average, the unemployment rates (the top two panels) of individuals without a college degree are more cyclically sensitive, rising by more in downturns and falling by more in expansions. At all times, the unemployment rates for those without a college degree are higher than the rates for those with a college degree.

The LFPR (the bottom panels) is lower for those with less education. Similar to the results by race and ethnicity, the LFPR exhibits little observable cyclical sensitivity. The gaps in the LFPR by educational attainment between those with a high school degree or less and the other two groups are large and persistent.

In his original paper, Okun noted that a high-pressure economy helps workers find employment and upskills the types of jobs they can obtain, translating into better wages, earnings, and household incomes. Figures 4 and 5 present analogous information with respect to real average hourly wages, annual own earnings (which accounts for both hourly earnings and hours of work), and annual household income. There is some cyclicality in all three measures, with all three rising faster in strong periods than in weak periods. That said, there is very little visual evidence that the strength of the labor market affects the gaps in these variables across less advantaged and more advantaged groups. In general, these aggregate income measures for blacks and Hispanics are far lower than the analogous measures for whites; similarly, the average incomes of those with lower educational attainment are well below those of persons with higher educational attainment.

Table 1. Gaps by Race and Ethnicity and Gender, Full Sample, Age 25–64 Years[a]

		Men		*Women*	
Characteristic	*Ethnicity*	*Constant*	Ugap	*Constant*	Ugap
Unemployment rate	Black	4.446***	0.909***	4.214***	0.513***
		(0.119)	(0.078)	(0.156)	(0.116)
	Hispanic	2.234***	0.394***	3.427***	0.339***
		(0.180)	(0.086)	(0.183)	(0.091)
Nonparticipation rate	Black	7.609***	0.077	–1.026**	0.081
		(0.170)	(0.128)	(0.440)	(0.247)
	Hispanic	–0.936***	–0.152	9.362***	–0.250*
		(0.296)	(0.152)	(0.358)	(0.132)
Hourly wages	Black	29.559***	–0.057	14.780***	–0.045
		(0.407)	(0.220)	(0.721)	(0.424)
	Hispanic	35.812***	–0.566	24.691***	–0.402
		(0.876)	(0.477)	(0.976)	(0.657)
Annual own earnings	Black	54.391***	1.163***	16.005***	2.286***
		(0.735)	(0.342)	(1.008)	(0.431)
	Hispanic	51.205***	0.634	46.906***	0.802*
		(1.505)	(0.585)	(1.203)	(0.436)
Household income	Black	37.497***	1.048**	52.804***	1.481***
		(1.074)	(0.485)	(1.354)	(0.420)
	Hispanic	39.516***	–0.077	43.747***	0.637
		(1.052)	(0.360)	(1.522)	(0.570)

Sources: Authors' estimates, using data from the U.S. Census Bureau, the Bureau of Labor Statistics (Current Population Survey), and the Congressional Budget Office (natural rate of unemployment).

a. Robust standard errors are in parentheses; $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$. Sample period is 1976:Q1–2018:Q4 for the employment-to-population ratio, unemployement rate, and labor force participation rate; 1987–2017 for annual own earnings and household income; and 1979:Q1–2018:Q4, when available, for hourly wages. The unemployment rate and nonparticipation rate gap for each group are defined as the outcome for the group indicated minus the outcome for the reference group. The wage, earnings, and income gaps for each group are defined as the outcome for the reference group minus the outcome for the group indicated. *Ugap* is defined as the aggregate unemployment rate minus the CBO's long-run natural rate of unemployment.

To document the greater cyclical sensitivity of the labor market and income experiences of less advantaged groups, on average, over the entire labor market cycle, tables 1 and 2 report estimates from a simple regression equation of this form:

$$y_{gt} = \alpha_0 + \alpha_1 * ugap_t + \varepsilon_t . \quad (1)$$

In table 1, the left-hand-side variable in each equation (denoted y_{gt} in equation 1) is the difference between a labor market– or income-related variable for the race and ethnicity and gender group (g) that is named in the line and column of the table, and the same variable for whites of the same

Table 2. Gaps by Education Level and Gender, Full Sample, Age 25–64 Years[a]

		Men		*Women*	
Characteristic	*Education level*	*Constant*	Ugap	*Constant*	Ugap
Unemployment rate	High school or less	3.350***	0.969***	3.291***	0.560***
		(0.106)	(0.052)	(0.068)	(0.038)
	Some college	1.556***	0.583***	1.509***	0.365***
		(0.038)	(0.019)	(0.051)	(0.047)
Nonparticipation rate	High school or less	9.848***	0.114	18.469***	0.179
		(0.231)	(0.119)	(0.324)	(0.146)
	Some college	3.715***	0.258	5.588***	0.237*
		(0.278)	(0.168)	(0.304)	(0.139)
Hourly wages	High school or less	53.694***	–0.264	58.512***	–0.535
		(1.629)	(1.117)	(1.279)	(0.910)
	Some college	33.728***	–0.213	35.725***	–0.290
		(1.386)	(0.927)	(1.351)	(0.893)
Annual own earnings	High school or less	88.480***	2.782**	97.156***	2.517***
		(3.290)	(1.103)	(1.847)	(0.643)
	Some college	54.065***	2.327**	50.452***	2.268***
		(3.036)	(0.955)	(1.802)	(0.668)
Household income	High school or less	69.102***	1.597*	77.731***	1.817***
		(2.416)	(0.793)	(1.429)	(0.557)
	Some college	42.519***	1.229*	43.705***	2.029***
		(1.957)	(0.631)	(1.632)	(0.567)

Sources: Authors' estimates, using data from the U.S. Census Bureau, the Bureau of Labor Statistics (Current Population Survey), and the Congressional Budget Office (natural rate of unemployment).

a. Robust standard errors are in parentheses; **p* <.10, ***p* < .05, ****p* < .01. Sample period is 1976:Q1–2018:Q4 for the unemployment rate and labor force participation rate; 1987–2017 for annual own earnings and household income; and 1979:Q1–2018:Q4, when available, for hourly wages. The unemployment rate and nonparticipation rate gap for each group are defined as the outcome for the group indicated minus the outcome for the reference group. The wage, earnings, and income gaps for each group are defined as the outcome for the reference group minus the outcome for the group indicated. *Ugap* is defined as the aggregate unemployment rate minus the CBO's long-run natural rate of unemployment.

gender. Thus, for example, the upper left block of coefficients pertains to a regression in which the left-hand-side variable is the unemployment rate for black men minus the unemployment rate for white men. Similarly, in table 2, the left-hand-side variable in each equation is constructed as the difference between a labor market– or income-related variable for the education and gender group that is named in the line and column of the table, and the same variable for individuals of the same gender and with a college degree or more. The regressions are run over the period 1976:Q1–2018:Q4. Importantly, to simplify the task of keeping track of signs, we define the nonparticipation rate as 1 minus the participation rate; similarly, for the earnings/income variables, we redefine the left-hand-side variable as 100 times the log of earnings/income for the reference group (for example, white women) minus the log of earnings/income for the comparison group (for example, black women). With this transformation, all the variables on the left-hand-side of regression equations are defined such that higher values represent worse outcomes, and a positive sign on the coefficient for *Ugap* indicates that the relatively disadvantaged group benefits more from each increment of labor market strengthening.

The coefficients of most interest to us in these tables are the ones that appear under the columns headed "*Ugap*." In the topmost block of results of table 1, the uniformly positive coefficients in these two columns replicate the finding of previous authors that, on average, when the labor market strengthens (that is, *Ugap* decreases), the unemployment rates for blacks and Hispanics decline by more than the unemployment rate for whites. Similarly, table 2 shows that the unemployment rates for individuals with a high school education or less and for individuals with some college education decline by more than the unemployment rate for individuals with a college degree or more. Moreover, in each of the tables, all eight of these slope coefficients are significantly different from zero at the 1 percent level.

In the blocks reporting results for the nonparticipation rate, a positive coefficient on *Ugap* indicates that as the labor market strengthens, the LFPR for the relatively marginalized group increases by more than the LFPR for the reference group—that is, the relatively marginalized group experiences a greater benefit as its relative nonparticipation rate falls. In this case, the slope coefficients are generally smaller in magnitude than they were for the unemployment rates and are of mixed sign and statistical significance—a result that may not be surprising, given the moderate cyclicality of this variable (Aaronson and others 2014). For blacks, the coefficients are positive but not statistically significant, while the two coefficients for Hispanics are negative (indicating that white participation has been more cyclically sensitive, on average, than has Hispanic participation). By educational attainment, all the coefficients are positive, though only statistically significant for women with some college at the 10 percent level.

The bottom three blocks of tables 1 and 2 report results for the three income-related measures that we examine (with the reminder that a positive slope coefficient is associated with the relatively disadvantaged group benefiting more from each increment of labor market strengthening). The gaps in average hourly earnings are not particularly cyclically sensitive; none of the four estimated slope coefficients shown in tables 1 and 2 is significantly different from zero, and all are negative. This result could reflect the changing composition of employment as the economy improves and more marginal workers with lower pay become employed (Daly and Hobijn 2017). It could also be that more of the relative improvement in labor income for less advantaged groups comes in the form of hours worked rather than hourly pay (Doniger 2019). Consistent with the latter-hypothesis, 15 of the 16 coefficients in the bottom two blocks (annual own earnings and annual household income) of tables 1 and 2 are positive, and 13 of these are significant at the 10 percent level or better.

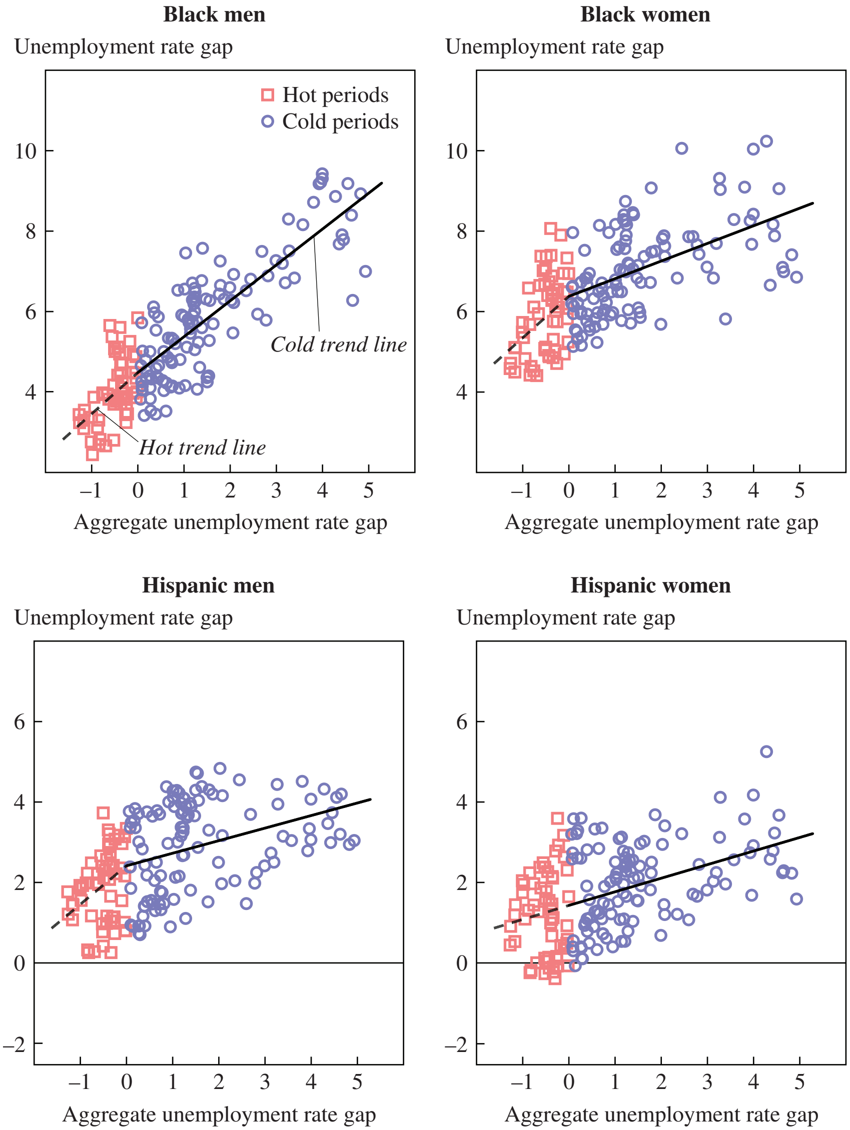
Overall, these results confirm those from previous studies, namely, that less advantaged groups experience a high-beta version of the cyclical sensitivity of labor market outcomes of more advantaged groups. Next, we consider whether that sensitivity differs significantly when the labor market is tight.

III.B. Are Hot Periods Different from Cold Periods?

To begin our examination of whether the average experience documented in tables 1 and 2 differs between hot and cold periods, figures 6 and 7 display scatter plots showing the differential unemployment experiences of our eight comparison groups relative to their white or more highly educated counterparts. In these figures, the variable plotted against the vertical axis is the difference between the unemployment rate for the comparison group relative to the unemployment rate for either whites or individuals with at least a college education; each differential variable is constructed separately for men and for women. The variable plotted against the horizontal axis is the aggregate unemployment rate gap; thus, observations further to the right in the figure come from periods when the labor market was looser (more slack) and observations further to the left come from periods when the labor market was tighter (less slack). To show average tendencies, we draw trend lines through the data points, noting that a flat line would indicate that the unemployment rate gap between the two groups is not sensitive to the tightness of the labor market. To ascertain whether the relative unemployment experience is different when the economy is operating in high-pressure mode, we allow each trend line to have a kink where the unemployment rate gap equals zero. If the responsiveness is the same in both hot and cold periods, the trend lines will be linear with no observable kink.

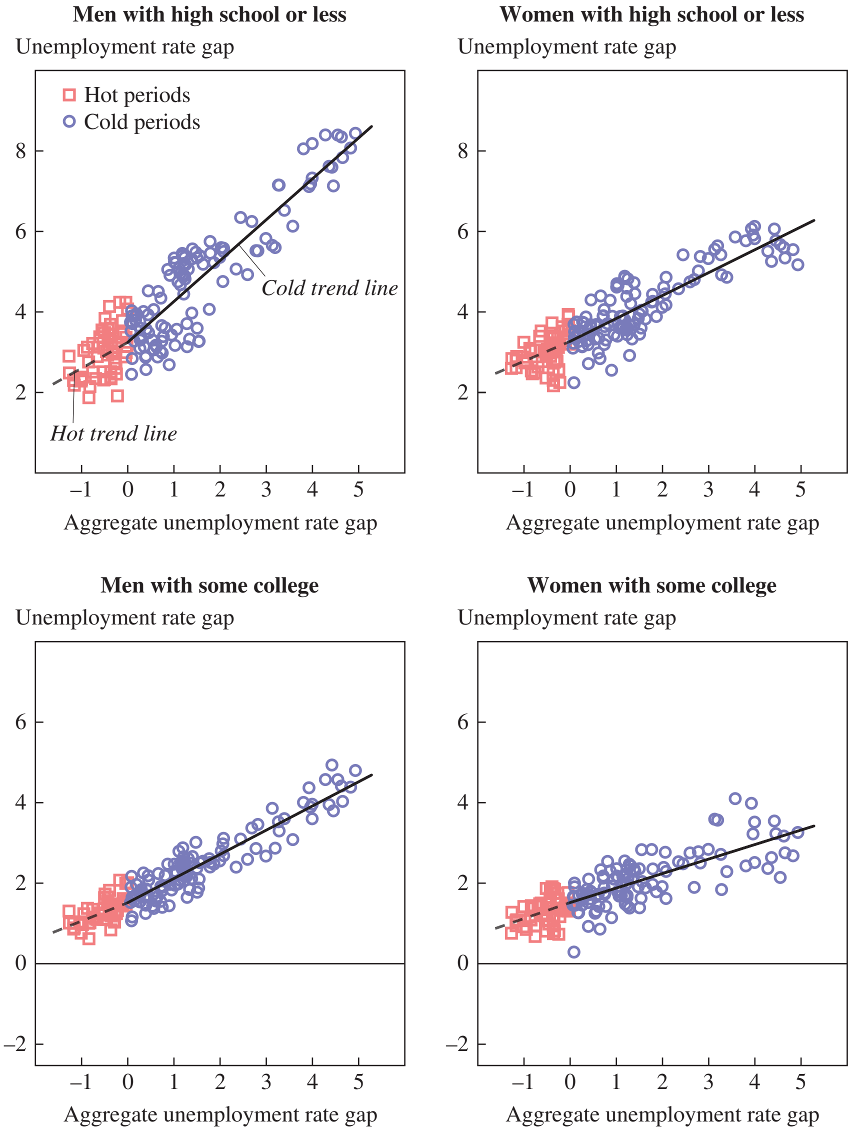
Figure 6 shows results for the unemployment rate by race and ethnicity. Pooling the roughly four decades in our sample, the lines are kinked downward for black women (the upper right panel) and Hispanic men (the bottom left panel), indicating that as the labor market moves into

Figure 6. Unemployment Rate Gap by Race and Ethnicity and by Gender, Age 25–64 Years



Sources: Authors' calculations, using data from U.S. Census Bureau and the Bureau of Labor Statistics (Current Population Survey).

Figure 7. Unemployment Rate Gap by Education Level and Gender, Age 25–64 Years



Sources: Authors' calculations, using data from U.S. Census Bureau and the Bureau of Labor Statistics (Current Population Survey).

high-pressure mode, not only do the unemployment rates of black women and Hispanic men continue to decline by more than the unemployment rate of their white counterparts, but the multiplier increases. In the econometric specification used to construct these panels, the process goes into reverse once the unemployment rate gap has reached its nadir. (Due to the limited number of data points, we did not test whether there was asymmetry depending on whether the economy was expanding or contracting.) As the unemployment rate comes back up toward its natural rate, the unemployment experience of black women and Hispanic men deteriorates more sharply than it does for their white counterparts, and by a wider margin than is estimated to occur once the unemployment rate moves above its natural rate. There is no discernible difference between hot and cold periods in the high-beta behavior of the unemployment rate of black men compared with white men, or for Hispanic women compared with white women.

Figure 7 compares the unemployment experience of individuals either with a high school degree or less, or with some college education, to that of individuals with a college degree or more. In no case is there evidence that hot periods are better for those with less than a college degree. In fact, as the aggregate unemployment rate moves below its natural rate, the unemployment rates for men either with a high school degree or less, or with some college, decline by less than they did earlier in the labor market cycle (indicated by the fact that the line is less steep to the left of $Ugap = 0$ than it is to the right). For women with a high school degree or less or some college education, hot and cold periods appear to differ little.

A natural question to ask is whether the basic relationships displayed in figures 6 and 7 have been stable over time. To answer this question, we divided our sample period into four labor market cycles—with each cycle defined as beginning in the quarter when the unemployment rate first exceeds the natural rate and ending in the quarter when the unemployment rate last falls below or equals the natural rate. We then conducted simple F tests to determine whether the null hypothesis of equality across the four slope coefficients can be rejected.[11] In the overwhelming majority of cases, the null hypothesis is rejected at the 5 percent level or better.

Tables 3 and 4 accordingly report coefficient estimates for regressions taking this form:

$$y_{gt} = \alpha_0 + \alpha_1 * ugap_t + \alpha_2 * hot\ dummy_t * ugap_t + \varepsilon_t \tag{2}$$

11. Throughout the paper, we conduct hypothesis tests using covariance matrices that are robust to serial correlation and heteroscedasticity.

where the regression is run separately for the sample as a whole and for each of the labor market cycles. As in equation 1, the left-hand-side variable in the regression is the difference between the unemployment rate for the comparison group, g, and that of their more advantaged counterparts (whites or those with a college education or more). The variable *hot dummy* takes a value of 1 when the overall unemployment rate is less than its natural rate and 0 otherwise.

The top row of table 3 reports results for the entire sample period taken as one—the same results as were shown in figure 6—while the remaining rows report results for each labor market cycle separately. Looking across the four cycles and the four race/ethnicity/gender pairs, in 15 of the 16 cases the trend line is estimated to have had a positive slope during cold periods (when $Ugap > 0$), confirming that these groups endured a high-beta version of the unemployment rate experience of their white counterparts.

Next, we turn to the question of whether that high-beta experience evolved once the labor market was tight. In a pattern that is repeated in later analyses, the relative improvement in the unemployment rates of black men and black and Hispanic women did not intensify during the high-pressure period of the late 1980s; this is reflected in the table by the fact that the estimated coefficients on the interaction term in these three cases are negative. However, in 10 of the other 12 cases (the exceptions being Hispanic men during the cycle of the early 2000s and Hispanic women during the current cycle), the coefficient on the interaction term is estimated to have been positive, meaning that the high-beta experience of the studied group intensified as the unemployment rate moved below its natural rate. In fact, in 6 of those 10 cases, the coefficient estimates suggest that the relative improvement when the labor market was tight was more than double the relative improvement when the labor market was slack. The coefficient on the interaction term is statistically significant and positive in 5 cases.

As shown in table 4, the results are somewhat weaker for the relative unemployment rates of groups stratified by educational attainment. The slope of the trend line in cold periods is estimated to have been positive in 15 of the 16 cycle-specific cases shown in the table. However, the increment to the slope during a hot labor market is of mixed sign, positive in 9 cycle-specific instances and negative the other 7 times. That said, the overall slope during high-pressure economies typically remained positive. Thus, though less educated individuals also undergo a high-beta version of the unemployment experience of those with at least a college education, there is little evidence that the beta has increased in hot labor markets, with

Table 3. Unemployment Rate Gaps by Race and Ethnicity, Gender, and Business Cycle, Age 25–64 Years[a]

	Men					
	Black			*Hispanic*		
Business cycle	*Slope when* Ugap > *0*	*Increment when* Ugap ≤ *0*	*Slope when* Ugap ≤ *0*	*Slope when* Ugap > *0*	*Increment when* Ugap ≤ *0*	*Slope when* Ugap ≤ *0*
All business cycles	0.881***	0.252	1.133	0.324***	0.566	0.890
	(0.102)	(0.347)		(0.110)	(0.481)	
1980:Q1–1990:Q3	0.854***	–0.426	0.428	0.272***	0.635	0.906
	(0.052)	(0.485)		(0.058)	(0.604)	
1990:Q4–2001:Q3	0.862***	0.193	1.055	0.678***	0.782***	1.460
	(0.121)	(0.307)		(0.124)	(0.288)	
2001:Q4–2007:Q4	0.254	0.511	0.765	0.871***	–0.660	0.211
	(0.407)	(1.234)		(0.243)	(0.584)	
2008:Q1–2018:Q4	0.905***	0.899*	1.804	0.501***	0.314	0.815
	(0.126)	(0.474)		(0.053)	(0.340)	

Sources: Authors' estimates, using data from the U.S. Census Bureau, the Bureau of Labor Statistics (Current Population Survey), and the Congressional Budget Office (natural rate of unemployment).

a. Robust standard errors are in parentheses; *p <.10, **p < .05, ***p < .01. The unemployment rate gap for each group is defined as the outcome for the group indicated minus the outcome for the reference group. *Ugap* is defined as the aggregate unemployment rate minus the CBO's long-run natural rate of unemployment.

Table 4. Unemployment Rate Gaps by Education Level, Gender, and Business Cycle, Age 25–64 Years[a]

	Men					
	High school or less			*Some college*		
Business cycle	*Slope when* Ugap > *0*	*Increment when* Ugap ≤ *0*	*Slope when* Ugap ≤ *0*	*Slope when* Ugap > *0*	*Increment when* Ugap ≤ *0*	*Slope when* Ugap ≤ *0*
All business cycles	0.985***	–0.206	0.779	0.591***	–0.087	0.504
	(0.063)	(0.251)		(0.025)	(0.105)	
1980:Q1–1990:Q3	1.003***	–0.358	0.645	0.534***	0.419	0.952
	(0.034)	(0.274)		(0.038)	(0.254)	
1990:Q4–2001:Q3	1.015***	0.031	1.046	0.594***	–0.151*	0.443
	(0.069)	(0.170)		(0.029)	(0.090)	
2001:Q4–2007:Q4	0.341**	–0.672*	–0.331	0.602***	–1.046**	–0.444
	(0.163)	(0.371)		(0.176)	(0.411)	
2008:Q1–2018:Q4	1.009***	0.053	1.062	0.569***	0.199	0.767
	(0.054)	(0.310)		(0.027)	(0.166)	

Sources: Authors' estimates, using data from the U.S. Census Bureau, the Bureau of Labor Statistics (Current Population Survey), and the Congressional Budget Office (natural rate of unemployment).

a. Robust standard errors are in parentheses; *p <.10, **p < .05, ***p < .01. The unemployment rate gap for each group is defined as the outcome for the group indicated minus the outcome for the reference group. *Ugap* is defined as the aggregate unemployment rate minus the CBO's long-run natural rate of unemployment.

Women					
Black			*Hispanic*		
Slope when Ugap > *0*	*Increment when* Ugap ≤ *0*	*Slope when* Ugap ≤ *0*	*Slope when* Ugap > *0*	*Increment when* Ugap ≤ *0*	*Slope when* Ugap ≤ *0*
0.445***	0.668	1.114	0.382***	–0.127	0.255
(0.143)	(0.427)		(0.126)	(0.515)	
0.555***	–0.308	0.247	0.725***	–1.819***	–1.094
(0.095)	(0.504)		(0.091)	(0.669)	
0.658***	0.171	0.828	–0.095	1.548***	1.453
(0.171)	(0.431)		(0.204)	(0.424)	
0.335	1.752*	2.087	1.101***	0.410	1.511
(0.357)	(0.866)		(0.211)	(0.516)	
0.443***	1.029***	1.472	0.518***	–0.024	0.494
(0.098)	(0.378)		(0.063)	(0.232)	

Women					
High school or less			*Some college*		
Slope when Ugap > *0*	*Increment when* Ugap ≤ *0*	*Slope when* Ugap ≤ *0*	*Slope when* Ugap > *0*	*Increment when* Ugap ≤ *0*	*Slope when* Ugap ≤ *0*
0.538***	0.039	0.577	0.337***	0.123	0.460
(0.046)	(0.155)		(0.059)	(0.136)	
0.469***	0.668	1.137	0.143***	0.702*	0.845
(0.051)	(0.497)		(0.049)	(0.377)	
0.501***	–0.077	0.424	0.459***	–0.119	0.340
(0.082)	(0.168)		(0.063)	(0.126)	
–0.169	1.327***	1.157	0.107	–0.556	–0.449
(0.185)	(0.449)		(0.125)	(0.340)	
0.520***	0.600***	1.119	0.354***	0.118	0.472
(0.064)	(0.216)		(0.068)	(0.221)	

Table 5. Increments to β When the Unemployment Rate Is below the Natural Rate[a]

Category	*Unemployment rate*	*Nonparticipation rate*
Black men	*	@
Black women	**	*
Hispanic men	*	
Hispanic women		@
Men with high school or less		*
Women with high school or less	**	**
Men with some college		
Women with some college		**

Sources: Authors' estimates, using data from the U.S. Census Bureau, the Bureau of Labor Statistics (Current Population Survey), and the Congressional Budget Office (natural rate of unemployment).

a. * At least three cycle-specific increments to β estimated to have been positive, of which no more than one is statistically significantly different from zero at the 10 percent level or better. ** At least two of the positive increments to β estimated to have been statistically significantly different from zero at the 10 percent level or better. @ Two cycle-specific increments to β estimated to have been positive and statistically significantly different from zero at the 10 percent level or better, but the other two increments estimated to have been negative.

the possible exception of women with a high school degree or less. We have estimated similar regressions for the nonparticipation rate, the results of which are available in the online appendix.

Table 5 provides a compact summary of the results from all these regressions. In the table, a single asterisk in a cell denotes that the estimated increment to β was positive in at least three of the four labor market cycles. A double asterisk adds the requirement that in at least two cases, positive increments were estimated to have been significantly different from zero at the 10 percent level of confidence or better. For completeness, we use an "@" sign to denote intermediate cases (four in number), in which two increments are estimated to have been positive and statistically significantly different from zero, but the other two increments were estimated to have been negative.

As can be seen in the first column of table 5, the results (as noted above) in the case of the unemployment rate are suggestive but not conclusive: Half of the cells in this column are blank, meaning that in those cases, either fewer than three of the estimated increments to β were positive or fewer than two were statistically significantly different from zero. In two of the eight cells, at least two increments were statistically significantly different from zero. In the nonparticipation column, six of the eight cells earn some form of marking—an interesting result, given that through most of the labor market cycle, the gaps in nonparticipation rates are noticeably less cyclical than are the gaps in unemployment rates. Nonetheless, our results suggest that once the labor market is operating in high-pressure mode, relatively marginalized persons are drawn into the labor market proportionately more than are relatively advantaged persons. Although this is not shown in the summary tables, the late 1990s seem to have brought widespread relative gains in participation rates: the increment to the slope during the hot period of that labor market cycle is positive for all racial and ethnic groups that we study, and these coefficients are statistically significant.

More generally, it is clear that labor market dynamics vary significantly across cycles, making it difficult to tell a simple story about the role of high-pressure economies. With that caveat, however, we read the evidence reported in table 5 as indicating that as the labor market has strengthened, the employment experiences of midlife African Americans and Hispanics age 25 to 64, as well as that of those with less than a college degree, have improved relatively more compared with whites and college-educated individuals of the same gender. Moreover, this observation holds true regardless of whether the labor market is operating in "cold" or "hot" territory. The evidence with respect to whether the relative experiences of disadvantaged groups have differed materially between cold and hot episodes is less clear, but leans in the direction of suggesting that there is a difference that skews in favor of these groups, particularly blacks and women with some college education or less. The relative improvement enjoyed by disadvantaged groups appears to have been particularly strong during the high-pressure labor market of the 1990s.[12]

III.C. Estimates with MSA Data

To test the robustness of these results, we use MSA-level data to look for evidence of the "high-beta" relationship between the labor market outcomes of disadvantaged groups and more advantaged groups and also for evidence that this relationship changes as the labor market

12. Although our assumption that the kink in the slope occurs when the unemployment gap is zero is intuitively appealing, in principle the kink could occur above or below that point. To assess this possibility, we also experimented with threshold specifications that allow the data to choose the point at which the kink occurs. For most groups, this version of the model chose a kink point that was between 1 and 2 percentage points above the natural rate; the exception was the unemployment differential for black men, for which the chosen kink point was ½ percentage point below the natural rate. For the unemployment and nonparticipation rate gaps, the slope coefficients during cold periods were similar to those shown in tables 3 and 4, despite the differences in the kink points. These specifications also tended to show an intensification of the high-beta experience for blacks and Hispanics below the chosen kink point (9 out of 12 cases for unemployment gaps, and 7 out of 12 cases for nonparticipation; we were unable to run this model for the 2001–7 period). And, as was the case for the specifications assuming a kink at $Ugap = 0$, the threshold results were weaker for relative unemployment gaps and nonparticipation gaps by educational attainment.

Table 6. Gaps by Demographic Group, Metropolitan Areas, Age 25–64 Years[a]

Characteristic	*Demographic group*	*Slope,* Ugap > *0*	*Increment,* Ugap < *0*
Unemployment rate	Black	0.476***	0.816**
		(0.172)	(0.394)
	Hispanic	0.305*	–0.238
		(0.171)	(0.341)
	High school or less	0.880***	0.246
		(0.104)	(0.201)
	Some college	0.477***	0.267**
		(0.078)	(0.133)
Nonparticipation rate	Black	0.326	1.054
		(0.252)	(0.832)
	Hispanic	–0.141	–0.745
		(0.312)	(0.803)
	High school or less	–0.0778	–0.268
		(0.165)	(0.436)
	Some college	–0.0533	0.701*
		(0.169)	(0.388)

Sources: Authors' estimates, using data from the U.S. Census Bureau and the Bureau of Labor Statistics (Current Population Survey).

a. Robust standard errors, clustered by metropolitan area, are in parentheses; *p <.10, **p < .05, ***p < .01. The unemployment rate and nonparticipation rate gap for each group are defined as the outcome for the group indicated minus the outcome for the reference group. All regressions include year and metropolitan-area fixed effects. Yearly data from 2004:Q3–2008:Q4 are used to calculate the natural rate of unemployment. *Ugap* is defined as the metropolitan-area unemployment rate minus the metropolitan-area natural rate of unemployment. Regressions then include 2009:Q1–2018:Q4. Regressions are weighted by population size. Metropolitan areas included have an average of 75 observations per demographic category and an average population of more than 500,000 over the 15-year period. Regressions on the black gap include 520 observations, on the Hispanic gap include 513 observations, on the high school or less gap include 530 observations, and on the some college gap include 540 observations.

enters a high-pressure period.[13] We define the natural rate in each metropolitan area as the average unemployment rate for that area for the period 2004:Q3–2008:Q4 and run the panel regression over the period 2009:Q1–2018:Q4, including year and metropolitan-area fixed effects.[14]

The results, shown in table 6, are consistent with the time-series analysis. The coefficients are of similar magnitude in absolute value and show some

13. This analysis is similar in spirit to those done by Kiley (2015), Leduc and Wilson (2019), Leduc and Wilson (2017), and Smith (2014)—all of whom use cross-metropolitan-area or cross-state variation to test the sensitivity of wage or price inflation to labor market slack.

14. Ideally, we would use a longer-length lag or some other filtering to compute the natural rate, but the time series of metropolitan-level data is not very long. As an alternative, we tried using a backward-looking, 7-year moving average of the unemployment rate. In this case, the coefficients on the unemployment rate gap are attenuated and statistically insignificant, likely because this measure puts too much weight on the high unemployment rates of the Great Recession in calculating the natural rate. The coefficients on the hot labor market interaction were more typically statistically significant in this specification.

Table 7. Increments to β When the Unemployment Rate Is below the Natural Rate[a]

Category	*Hourly wages*	*Annual own earnings*	*Household income*
Black men			
Black women		*	
Hispanic men			
Hispanic women		*	*
Men with high school or less			
Women with high school or less		*	
Men with some college			
Women with some college		*	

Sources: Authors' estimates, using data from the U.S. Census Bureau, the Bureau of Labor Statistics (Current Population Survey), and the Congressional Budget Office (natural rate of unemployment).

a. * For hourly wages, at least three cycle-specific increments to β estimated to have been positive, of which no more than one is statistically significantly different from zero at the 10 percent level or better. For annual own earnings and household income, estimated increment to β is positive but not significantly different from zero. ** For hourly wages, at least two of the positive increments to β estimated to have been statistically significantly different from zero at the 10 percent level or better. For annual own earnings and household income, estimated increment to β is positive and statistically significantly different from zero at the 10 percent level or better. @ For hourly wages, two cycle-specific increments to β estimated to have been positive and statistically significantly different from zero at the 10 percent level or better, but the other two increments estimated to have been negative. Not relevant for annual own earnings or household income.

evidence that high-pressure economies are particularly beneficial for disadvantaged groups. For example, the unemployment rates of the disadvantaged groups are more cyclical, and this relationship is statistically significant. Moreover, during the high-pressure phase of the cycle, this relationship appears to intensify for all groups except Hispanics, and it is statistically significant for blacks and those with some college education. With regard to the nonparticipation rate, the results using the metropolitan-level data are weaker—the slope coefficient in cold periods is positive only for blacks, and even then it is not statistically significant. When the economy is in a high-pressure state, the evidence suggests that the participation rate gap closes by more for blacks and for those with some college education, but it is only statistically significant for the latter group.[15]

III.D. Earnings and Income

Table 7 provides a scoring of results for the three relative income variables that we inspect, based on average hourly wages, annual own earnings,

15. We note two caveats to this analysis. First, we do not break out men and women separately, so the results cannot speak to the differences by gender that are evident in the time-series analysis (for instance, the high cyclicality of the employment-to-population ratio for Hispanic men and black women). Second, the data used for this analysis are all from the final labor market cycle of our time-series analysis.

and annual household income. For average hourly wages, we use the same method that we used to construct the scoring reported in table 5. For the own earnings and household income variables, we use a simpler method because the underlying data are annual: We award one asterisk if the estimated coefficient (by construction, over the whole sample period) is positive, and two asterisks if it is significantly so.[16]

The contrast between tables 5 and 7 is plain: Whereas a slight majority of cells in table 5 showed some marking, the great majority of cells in table 7 are blank, signifying that when the labor market is tight, β generally does not shift in a manner that is favorable to the relatively marginalized group. Results shown in the online appendix go a step further and demonstrate that, in fact, relative income gaps actually widen in about half the 24 cases that we examine (8 demographic pairs and 3 relative income variables).

The results on earnings gaps are broadly consistent with previous research that finds lower wage cyclicality among less advantaged groups than among more advantaged groups. For less advantaged workers, institutional constraints such as the minimum wage are more likely to bind in cold periods (Hoynes 2000); and in hot periods, more advantaged workers with higher skills are more likely to see rapid wage increases (Daly and Hobijn 2017; Doniger 2019). In terms of household earnings and income, previous research has shown that families smooth through income variability, including variability induced by unemployment rate shocks, using the social safety net and changes to family labor supply (Dynarksi and Gruber 1997). This behavior puts a floor under families in cold periods. In hot periods, the relatively larger wage gains going to more advantaged workers are likely amplified by patterns of household formation that result in the presence of multiple advantaged workers in the same household (Eika, Mogstad, and Zafar 2018). To sum up, in a hot economy, less advantaged groups improve relative to more advantaged groups in their employment experiences; in contrast, more advantaged groups experience relatively larger gains in hourly wages and income. Future research linking these findings to broader implications for economic welfare is needed.

III.E. Results for Individuals between the Age of 16 and 24 Years

Okun's hypothesis particularly focused on the advantage of hot labor markets to young workers, and indeed, the labor market experience of

16. Recall that for the earnings and income variables, we define the gaps as the earnings or income level for whites or college graduates relative to that for the indicated group, so that a positive coefficient signifies a narrowing of the gap as the unemployment gap declines.

individuals at the lower end of the age spectrum may differ importantly from the labor market experience of people age 25 to 64. To ascertain whether differences across age groups are important, we briefly review results that are analogous to those we have already shown for those age 25 to 64, but in this case for people between the age of 16 and 24.

Table 8 presents the relative cyclical sensitivities of the unemployment rate gaps of young adults for each of the four demographic pairs in our focus, in the same format as table 3. For African Americans, these results are reasonably straightforward to characterize. In all the episodes we considered, the unemployment rates of young African Americans were more cyclically sensitive than the unemployment rates of their white counterparts, and they became even more so as the unemployment rate moved below the CBO's natural rate. (This result is signified by the fact that all eight cycle-specific point estimates reported in the first and second columns for African American men and women are positive.) Looking across age groups, the fact that the point estimates are generally larger, in absolute value, than the point estimates in table 3 shows that young blacks also experience more relative cyclical variation in their unemployment rates (relative to their white counterparts) than do midlife blacks.

For young Hispanics, the results are a little more uneven. Young Hispanic men exhibit greater cyclicality in their unemployment rates in all four labor market cycles, while young Hispanic women exhibit greater cyclicality in unemployment rates in three of the four. The evidence regarding the question of whether the benefits of a strengthening labor market skew more in favor of young Hispanics relative to whites once the economy is operating in high-pressure mode is mixed. Of the eight cycle-specific interaction coefficients for young Hispanic men and women, only five are positive (only two of which are statistically significant).

III.F. Urban versus Rural Differences

We examine one final divide of interest: the difference in economic performance between more and less urbanized areas, or what the CPS denotes metropolitan and nonmetropolitan areas.[17] Alison Weingarden (2017) has documented that labor force participation rates in nonmetropolitan areas have decreased relative to those in metropolitan areas, going

17. Metropolitan areas are those that contain a significant population nucleus, of at least 50,000 people, and adjacent communities that have a high degree of integration with that nucleus. Nonmetropolitan areas are the complement. Strictly speaking, they are not synonymous with rural areas.

Table 8. Unemployment Rate Gaps by Race and Ethnicity, Gender, and Business Cycle, Age 16–24 Years[a]

	Men					
	Black			*Hispanic*		
Business cycle	*Slope when* Ugap > *0*	*Increment when* Ugap ≤ *0*	*Slope when* Ugap ≤ *0*	*Slope when* Ugap > *0*	*Increment when* Ugap ≤ *0*	*Slope when* Ugap ≤ *0*
All business cycles	0.872** (0.355)	1.046 (1.179)	1.918	0.500*** (0.164)	–0.102 (0.653)	0.398
1980:Q1–1990:Q3	1.470*** (0.226)	2.676 (1.678)	4.146	0.675*** (0.144)	–0.041 (1.353)	0.634
1990:Q4–2001:Q3	1.123* (0.646)	0.446 (1.366)	1.569	0.184 (0.344)	1.412** (0.695)	1.596
2001:Q4–2007:Q4	0.272 (0.814)	6.352*** (2.034)	6.624	1.065 (0.705)	0.147 (2.097)	1.212
2008:Q1–2018:Q4	1.160*** (0.172)	2.311*** (0.779)	3.471	0.533*** (0.087)	–0.414 (0.516)	0.119

Sources: Authors' estimates, using data from the U.S. Census Bureau, the Bureau of Labor Statistics (Current Population Survey), and the Congressional Budget Office (natural rate of unemployment).

a. Robust standard errors are in parentheses; *p <.10, **p < .05, ***p < .01. The unemployment rate gap for each group is defined as the outcome for the group indicated minus the outcome for the reference group. *Ugap* is defined as the aggregate unemployment rate minus the CBO's long-run natural rate of unemployment.

back at least a decade. More recently, the improvement in the unemployment rate has lagged in nonmetropolitan areas, with the result that employment rates in these areas have fallen further behind those of metropolitan areas.

That said, the difference in labor market outcomes across metro and nonmetro areas seems to be mostly structural and does not appear to be particularly sensitive to the business cycle. For instance, as can be seen in the top panel of figure 8, the unemployment rates in metro and nonmetro areas are very similar, both in terms of their levels and cyclical amplitudes.[18] In fact, the data indicate that the unemployment rate in metro areas is a little more cyclically sensitive than the unemployment rate in nonmetro areas. In contrast, the participation rates are not particularly cyclical. When, as shown in table 9, we regress the difference in the unemployment rate or labor force participation rate (nonmetro minus metro) on the aggregate unemployment rate gap and a hot labor market interaction, all the coefficients are close to zero. Moreover, the coefficient on the unemployment rate gap, which is statistically significant, is the opposite of what one would

18. An exception to the typically tight co-movement was the period of the 1980s, when rural areas were devastated by a farm crisis (Barnett 2000).

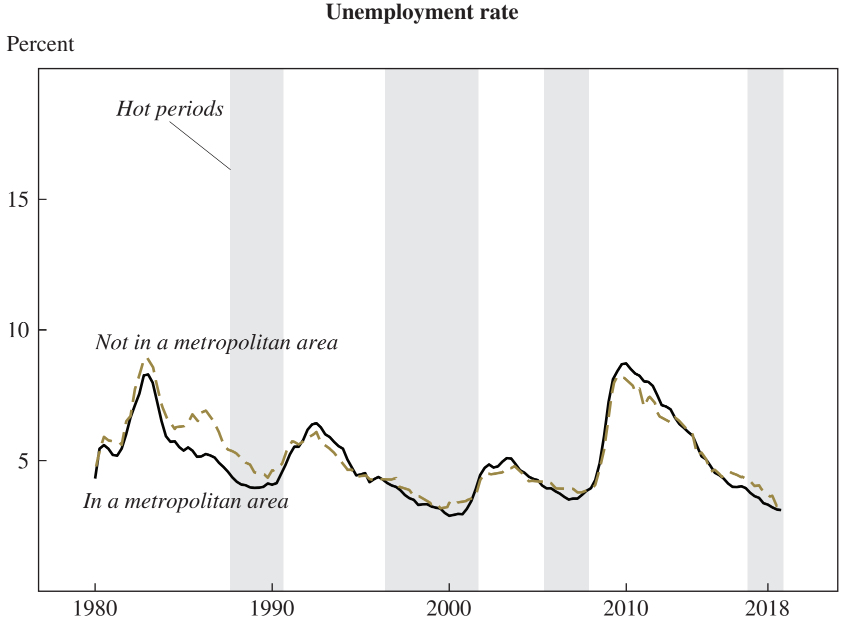
Women					
Black			*Hispanic*		
Slope when Ugap >*0*	*Increment when* Ugap ≤ *0*	*Slope when* Ugap ≤ *0*	*Slope when* Ugap > *0*	*Increment when* Ugap ≤ *0*	*Slope when* Ugap ≤ *0*
0.871	3.280*	4.151	0.483**	–0.394	0.089
(0.573)	(1.783)		(0.186)	(0.922)	
1.401***	5.849***	7.250	0.749***	–2.664***	–1.915
(0.223)	(1.996)		(0.149)	(0.901)	
1.598***	1.923**	3.521	–0.635	2.873**	2.238
(0.383)	(0.722)		(0.810)	(1.304)	
1.567*	1.347	2.914	0.973	1.901	2.873
(0.870)	(2.313)		(0.659)	(1.254)	
1.256***	3.789***	5.045	0.734***	1.191	1.924
(0.194)	(1.029)		(0.107)	(0.748)	

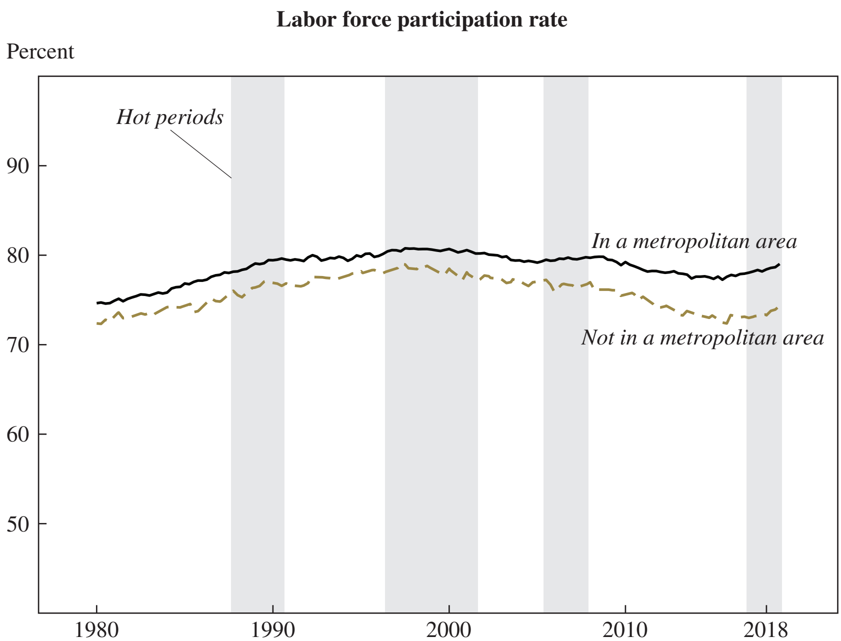
expect if economic expansions were bringing rural area outcomes closer to those in metro areas. Furthermore, there is no evidence that the relationship changes when the unemployment rate falls below its natural rate. These results do not change if we distinguish between small and large metro areas (not shown). Hence, though the evidence is clear that rural and to a lesser extent small metro area labor markets are falling behind those in larger metropolitan areas, the causes seem to be structural and are not ameliorated by a strong national labor market.

III.G. Hysteresis

Overall, it is clear that, as the aggregate labor market strengthens, disadvantaged workers benefit disproportionately, and there is suggestive evidence that this high-beta experience intensifies when the labor market is especially strong. Moreover, in Okun's original conception, high-pressure economies have an additional impact, because an individual who becomes employed may gain skills and networks that improve future employment prospects. To the extent that this dynamic exists, gains that start out as a result of the strong state of the business cycle could end up having beneficial longer-term effects on individual outcomes—what has been called positive

Figure 8. Labor Force Statistics by Metropolitan Area Status, Age 25–64 Years, 1980–2018





Sources: U.S. Census Bureau; Bureau of Labor Statistics (Current Population Survey).

Table 9. Nonmetropolitan/Metropolitan Gaps, 1980:Q1–2018:Q3, Age 25–64 Years[a]

Variable	*Slope,* Ugap > *0*	*Increment,* Ugap < *0*
Unemployment rate	–0.114*	0.106
	(0.062)	(0.204)
Nonparticipation rate	0.0534	–0.158
	(0.104)	(0.484)
Observations	156	

Sources: Authors' estimates, using data from the U.S. Census Bureau and the Bureau of Labor Statistics (Current Population Survey).

a. Newey–West standard errors are in parentheses; *p <.10, **p < .05, *** p < .01. The unemployment rate and nonparticipation rate gap are defined as the outcome for nonmetropolitan areas minus the outcome for metropolitan areas. *Ugap* is defined as the aggregate unemployment rate minus the CBO's long-run natural rate of unemployment.

hysteresis. Moreover, if these individual outcomes result in improvements in the economy overall—for instance, a lower unemployment rate on average or higher trend labor force participation—this would also boost the economy's potential growth rate.

Our approach to this question follows the strand in the literature that has looked for evidence of hysteresis in the aggregate data. Olivier Blanchard and Lawrence Summers (1986) describe hysteresis as the dependence of the current rate of employment on past realizations, and they find evidence of it in Europe, but little in the United States. As noted by Magnus Gustavsson and Pär Österholm (2007), in the macroeconomics literature, hysteresis has generally been interpreted as being reflected in the existence of a unit root in the unemployment rate. The evidence on this has, however, been mixed. Frank Song and Yangru Wu (1997) and Gustavsson and Österholm (2007) find little evidence of a unit root in unemployment in the United States. A few studies have also looked for evidence of a unit root in the employment-to-population ratio. Theoretically, this makes sense, because, as we have shown above, individuals adjust along the participation rate margin as well as the unemployment rate margin over the course of the business cycle. And indeed, the evidence for a unit root in the employment-to-population ratio seems a bit stronger (Gustavsson and Österholm 2007; Fallick and Krowlikowski 2018).

Here we repeat this time-series exploration of the question, updating past analysis to include data from the time of the Great Recession and through the current expansion. In addition, we examine unemployment and (non)participation rates by race and ethnicity and by level of education to explore the possibility that, even if aggregate statistics do not show clear evidence of hysteresis, it may be apparent in the labor market outcomes

Table 10. Univariate Unit-Root Tests, Age 25–54 Years[a]

Variable	*DF–GLS*	*Zivot–Andrews*	*Lags*
Unemployment rate	–3.164**	–5.051*	9, 3
Nonparticipation rate	–1.798	–3.548	10, 3

Sources: Authors' estimates, using data from the U.S. Census Bureau and the Bureau of Labor Statistics (Current Population Survey).

a. For DF–GLS, the lag is determined by the Ng–Perron test with generalized least squares. For Zivot–Andrews, an endogenously determined break is allowed in the intercept and trend, the lag is determined by Akaike's Information Criterion, and the 5 percent critical value is –5.08; $*p < .10$, $**p < .05$, $***p < .01$.

of specific groups. It is also important to note that the identification for this exercise comes from the entire sample, not just periods in which there are high-pressure economies, and so we do not distinguish the presence of positive versus negative hysteresis. As in our previous analysis, the tests are done using quarterly data from the CPS; however, because the aging of the population imparts a trend to the aggregate participation rate that could confound the results, we focus on the population age 25–54.

One of the problems with identification of a unit root is that if the data follow a trend or have a break, this can result in a spurious failure to reject a unit root. Indeed, a further inspection of figure 1 shows the unemployment rate drifting down between the 1980s and early 2000s, a time when some evidence suggests that the natural rate was falling, at least in part due to the aging of the baby boomers (Barnichon and Mester 2018; Staiger, Stock, and Watson 2001). The labor force participation rate more clearly has an uptrend, driven largely by the rapid increase in women's labor force participation, but there appears to be a break in that uptrend starting in the mid-1990s. For this reason, we select for our analysis tests that allow us to control for these trends and that include lags to eliminate serial correlation in the errors: the augmented Dickey–Fuller test with generalized least squares detrending and the Zivot–Andrews test, which allows for the possibility of breaks in the intercept and trend, with the break points determined endogenously. Both these tests have the null hypothesis that the series has a unit root.

As can be seen in table 10, the tests indicate that the unemployment rate lacks a unit root, consistent with the previous literature on the topic. In contrast, the tests do not reject that the labor force participation rate has a unit root. Table 11 shows the results for variables broken out by race and gender. The existence of a unit root in the unemployment rate is clearly rejected for white and black men and for Hispanic women. In contrast, the tests fail to reject a unit root for white women, suggesting hysteresis. For

Table 11. Univariate Unit-Root Tests, Age 25–54 Years, by Race/Ethnicity and Gender[a]

Characteristic	*Race/ethnicity*	*Gender*	*DF–GLS*	*Zivot–Andrews*	*Lags*
Unemployment rate	White	Men	–2.985**	–5.616***	13, 2
		Women	–2.466	–4.800	13, 2
	Black	Men	–3.358***	–5.024*	10, 3
		Women	–3.009**	–4.194	6, 3
	Hispanic	Men	–2.735*	–4.466	12, 3
		Women	–3.464**	–5.069*	13, 3
Nonparticipation rate	White	Men	–1.827	–3.164	13, 0
		Women	–2.268	–4.897*	13, 2
	Black	Men	–1.975	–3.719	7, 2
		Women	–1.258	–3.778	12, 3
	Hispanic	Men	–1.867	–4.399	12, 2
		Women	–0.973	–3.790	10, 3

Sources: Authors' estimates, using data from the U.S. Census Bureau and the Bureau of Labor Statistics (Current Population Survey).

a. For DF–GLS, the lag is determined by the Ng–Perron test with generalized least squares. For Zivot–Andrews, an endogenously determined break is allowed in the intercept and trend, the lag is determined by Akaike's Information Criterion, and the 5 percent critical value is –5.08; *p <.10, **p < .05, ***p < .01.

black women and Hispanic men, the results are inconclusive. With respect to the nonparticipation rate, the tests indicate the presence of a unit root for each of the groups defined by race, ethnicity, and gender.

Table 12 provides an assessment of the evidence of hysteresis for different education groups. The results clearly reject the presence of a unit root in the unemployment rate for men and women with a college education. For the remaining groups, the tests are less conclusive—with one of the tests rejecting the unit root. The tests almost unanimously fail to reject a unit root in the nonparticipation rate for men and women at all levels of education.[19]

These findings are consistent with there being positive spillovers from an expansion that could have lasting benefits for individuals and the economy, particularly along the participation rate margin, because the tests were consistent with hysteresis in the participation rate for nearly all groups. That said, one caveat to the analysis is that the microeconometric literature on hysteresis, which primarily focuses on the potentially lasting damage of recessions, suggests that employment gains are not expected to

19. We performed several robustness tests. Because a number of studies have suggested that the severity of the Great Recession may have led to an unusual degree of negative hysteresis (Yagan, forthcoming), we reran the tests on a sample ending in 2007:Q4, but the results were similar. Using the log odds ratio instead of the rate in order to avoid the problem that the rates are bounded between 0 and 1 also did not materially change the results.

Table 12. Univariate Unit-Root Tests, Age 25–54 Years, by Education Level and Gender[a]

Characteristic	*Education level*	*Gender*	*DF–GLS*	*Zivot–Andrews*	*Lags*
Unemployment rate	High school or less	Men	–2.793*	–5.352**	10, 2
		Women	–2.396	–5.397**	12, 3
	Some college	Men	–2.465	–5.769***	12, 3
		Women	–2.217	–4.890*	12, 3
	College or more	Men	–2.928**	–5.995***	13, 3
		Women	–2.694*	–4.879*	8, 0
Nonparticipation rate	High school or less	Men	–2.873*	–2.446	7, 2
		Women	–1.439	–3.330	10, 2
	Some college	Men	–1.553	–4.081	8, 1
		Women	–2.100	–4.244	11, 2
	College or more	Men	–1.638	–4.523	8, 3
		Women	–1.802	–4.289	13, 2

Sources: Authors' estimates, using data from the U.S. Census Bureau and the Bureau of Labor Statistics (Current Population Survey).

a. For DF–GLS, the lag is determined by the Ng–Perron test with generalized least squares. For Zivot–Andrews, an endogenously determined break is allowed in the intercept and trend, the lag is determined by Akaike's Information Criterion, and the 5 percent critical value is –5.08; $*p < .10$, $**p < .05$, $***p < .01$.

be long-lived (Hotchkiss and Moore 2018; Kahn 2010; Kondo 2015; and Oreopolous and others 2012).[20]

IV. The Potential Costs of a High-Pressure Economy

We have thus far focused on potential benefits of a high-pressure economy. However, running a hot economy also brings with it potential costs that policymakers should take into account.

Perhaps the most obvious risk associated with tight labor markets is the possibility of an unwelcome rise in inflation. Such a concern may seem unwarranted at present, given the apparent flattening of the Phillips curve in recent years, along with the observations that inflation has consistently run below the Federal Reserve's target for many of the past six years and that inflation expectations appear to be well anchored (see figure 1). However, it is worth remembering that the last time the unemployment rate was this low—in the late 1960s—inflation (as measured by the Personal Consumption Expenditures Price Index) moved up from less than 2 percent in 1965 to nearly 5 percent by 1970. In particular, policymakers at the time

20. In contrast, these studies find the impact of macroeconomic conditions on wages tends to last longer (also see Hagedorn and Manovskii 2013).

judged that an unemployment rate of about 4 percent was sustainable in the longer run (Orphanides and Williams 2013). In retrospect, however, the CBO now estimates the natural rate of unemployment to have been between 5½ and 6 percent in the second half of the 1960s. Moreover, a flatter Phillips curve may not be an unalloyed benefit: If inflation were somehow to become anchored at some level well above the FOMC's preferred level and the Phillips curve were to remain flat, the cost of bringing inflation down might be very high in terms of lost employment and output.

A second risk of a high-pressure economy, also macroeconomic in nature, has to do with the possibility of excessive risk-taking in financial markets and a resulting destabilization of the financial system. Again, current circumstances do not suggest that this is an imminent risk. For example, although the Federal Reserve's (2019a) latest *Financial Stability Report* characterizes valuation pressures as somewhat elevated, the report also notes that large banks are strongly capitalized and concludes that funding risks in the financial system are low relative to the period leading up to the financial crisis. That said, the most recent two recessions were precipitated by financial imbalances that were difficult to identify in real time. Also, some other observers are less sanguine. Of particular note, the Bank for International Settlements' (2018) *Annual Economic Report* expresses the concern that the accommodative stance of monetary policy that has helped to sustain the expansion and contributed to record-low unemployment has also resulted in building financial vulnerabilities—including a sustained rise in global debt-GDP ratios—that have increased the fragility of the economy.

Third, a hot economy has the potential to distort incentives, leading to decisions that emphasize short-run economic gains at the cost of longer-run sustainable economic progress. One example is the decision by younger individuals as to whether they should work or enroll in school. From a theoretical standpoint, schooling decisions may be influenced by the opportunity cost of attending school and by the direct financial costs of attendance, both of which may vary over the business cycle (though in opposite directions).[21] However, the empirical evidence indicates that enrollment rates tend to be countercyclical, suggesting that the short-term benefits of a high-pressure economy may hinder the building of sustainable career opportunities by incentivizing young people to drop out of school at a critical point in their academic career or to take an unstable job

21. See, for example, Dellas and Sakellaris (2003).

that may disappear with the next recession, rather than invest in training opportunities.[22]

Similarly, a high-pressure economy may encourage firms to focus on short-term economic profits at the expense of decisions aimed at enhancing their longer-run viability. For example, the owners of a firm may decide to defer maintenance of machinery, reorganizations, or research-and-development activities in a strong economy because the cost of potential forgone sales is viewed as too high. If so, the firm's future productivity may suffer as a result. More broadly, a high-pressure economy can potentially hinder the reallocation of resources from more productive to less productive activities by reducing the pressures on less productive firms to close down.[23]

V. Conclusions

So where do we stand? A few observations seem clear. First, as previous researchers have shown, when the economy weakens, everyone suffers; and when the economy strengthens, everyone benefits. This is seen most clearly in unemployment rates: Over our entire sample, the unemployment rates of each group we study move in tandem with the aggregate unemployment rate. Second, like others, we also find that the fluctuations of less advantaged groups—including blacks, Hispanics, and those with less than a college education—are more pronounced. When the labor market weakens, these groups tend to suffer disproportionately; when it recovers, their experience improves disproportionately. Third, inspired by Arthur Okun, we have also searched for evidence that high-pressure economies are qualitatively different, and we have found suggestive evidence that this is the case. A high-pressure economy does afford greater improvement for some less advantaged groups—most notably blacks and women with less than a college degree—in some key labor market variables, although the evidence is complicated by the heterogeneity observed across the various cycles. Finally, we also find suggestive evidence that these benefits persist

22. For evidence on four-year college enrollment, see Dellas and Sakellaris (2003). For evidence on enrollment at community colleges, see Betts and McFarland (1995). For evidence on high school enrollment, see Dellas and Koubi (2003).

23. Research on this issue focuses mostly on the behavior of firms in recessions rather than in strong economies. See, for example, Hall (1991), Caballero and Hammour (1994), and Aghion and Howitt (1992). Aghion and Saint-Paul (1998) and Legrand and Hagemann (2017) provide good overviews of both mechanisms.

at least for a while, particularly along the dimension of the labor force participation rate. All in all, the evidence presented here supports the idea that high-pressure economies are different than normal expansions—but just how different remains a topic for further study.

ACKNOWLEDGMENTS We are grateful to Francisca Alba, Neil Gerstein, Bo Yeon Jang, and Morgan Smith for excellent research assistance and to Tomaz Cajner and Chris Nekarda for providing already-constructed data sets and data dictionaries for some of the CPS data used in our analyses.

References

Aaronson, Stephanie, Tomaz Cajner, Bruce Fallick, Felix Galbis-Reig, Christopher Smith, and William Wascher. 2014. "Labor Force Participation: Recent Developments and Future Prospects." *Brookings Papers on Economic Activity*, Fall, 197–255. https://www.brookings.edu/wp-content/uploads/2016/07/Fall2014BPEA_Aaronson_et_al.pdf.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth through Creative Destruction." *Econometrica* 60, no. 2: 323–51.

Aghion, Philippe, and Gilles Saint-Paul. 1998. "Virtues of Bad Times: Interaction between Productivity Growth and Economic Fluctuations." *Macroeconomic Dynamics* 2, no. 3: 322–44.

Akerlof, George A., Andrew K. Rose, and Janet L. Yellen. 1988. "Job Switching and Job Satisfaction in the U.S. Labor Market." *Brookings Papers on Economic Activity*, no. 2: 495–594. https://www.brookings.edu/wp-content/uploads/1988/06/1988b_bpea_akerlof_rose_yellen_ball_hall.pdf.

Bank for International Settlements. 2018. *Annual Economic Report 2018*. Basel: Bank for International Settlements. https://www.bis.org/publ/arpdf/ar2018e.htm.

Barnett, B. J. 2000. "The U.S. Farm Financial Crisis of the 1980s." *Agricultural History* 74: 366–80.

Barnichon, Regis, and Geert Mesters. 2018. "On the Demographic Adjustment of Unemployment." *Review of Economics and Statistics* 100, no. 2: 219–31.

Betts, Julian R., and Laurel L. McFarland. 1995. "Safe Port in a Storm: The Impact of Labor Market Conditions on Community College Enrollments." *Journal of Human Resources* 30, no. 4: 741–65.

Blanchard, Olivier, and Lawrence Summers. 1986. "Hysteresis and the European Unemployment Problem." In *NBER Macroeconomics Annual 1986*, edited by Stanley Fischer. Cambridge, Mass.: MIT Press.

Bradbury, Katherine. 2000. "Rising Tide in the Labor Market: To What Degree Do Expansions Benefit the Disadvantaged?" *New England Economic Review*, May–June, 3–33.

Caballero, Ricardo J., and Mohamad L. Hammour. 1994. "The Cleansing Effect of Recessions." *American Economic Review* 84. no. 5: 1350–68.

Cajner, Tomaz, Tyler Radler, David Ratner, and Ivan Vidangos. 2017. "Racial Gaps in Labor Market Outcomes in the Last Four Decades and over the Business Cycle." Finance and Economics Discussion Series 2017-071. Washington: Board of Governors of the Federal Reserve System. https://doi.org/10.17016/FEDS.2017.071.

Dal Pont Legrand, Muriel, and Harald Hagemann. 2017 "Retrospectives: Do Productive Recessions Show the Recuperative Powers of Capitalism? Schumpeter's Analysis of the Cleansing Effect." *Journal of Economic Perspectives* 31, no. 1: 245–56.

Daly, Mary C., and Bart Hobijn. 2017. "Composition and Aggregate Real Wage Growth." *American Economic Review (Papers and Proceedings)* 107, no. 5: 349–52.

Daly, Mary C., Bart Hobijn, and Joseph Pedtke. 2019. "Labor Market Dynamics and Black–White Earnings Gaps." FRBSF Working Paper. San Francisco: Federal Reserve Bank of San Francisco.

Dellas, Harris, and Vally Koubi. 2003. "Business Cycles and Schooling." *European Journal of Political Economy* 19, no. 4: 843–59.

Dellas, Harris, and Plutarchos Sakellaris. 2003. "On the Cyclicality of Schooling: Theory and Evidence." *Oxford Economic Papers* 55, no. 1: 148–72.

Devereux, Paul. 2002. "Occupational Upgrading and the Business Cycle." *Labour* 16, no. 3: 423–52.

Doniger, Cynthia L. 2019. "Do Greasy Wheels Curb Inequality?" Finance and Economics Discussion Series, Federal Reserve Board, Washington.

Dynarski, Susan, and Jonathan Gruber. 1997. "Can Families Smooth Variable Earnings?" *Brookings Papers on Economic Activity*, Spring, 229–84. https://www.brookings.edu/wp-content/uploads/1997/01/1997a_bpea_dynarski_gruber_moffitt_burtless.pdf.

Eika, Lasse, Magne Mogstad, and Basit Zafar. 2018. "Educational Assortative Mating and Household Income Inequality." NBER Working Paper 20271. Cambridge, Mass.: National Bureau of Economic Research.

Fallick, Bruce, and Pawel Krolikowski, 2018. "Hysteresis in Employment among Disadvantaged Workers," Federal Reserve Bank of Cleveland, Working Paper no. 18-01.

Federal Reserve Board of Governors. 2019a. *Financial Stability Report: May 2019*. Washington: Federal Reserve Board of Governors. https://www.federalreserve.gov/publications/files/financial-stability-report-201905.pdf.

———. 2019b. "Review of Monetary Policy Strategy, Tools, and Communications." https://www.federalreserve.gov/monetarypolicy/review-of-monetary-policy-strategy-tools-and-communications-fed-listens-events.htm.

FOMC (Federal Open Market Committee). 2019a. "Economic Projections of Federal Reserve Board Members and Federal Reserve Bank Presidents under Their Individual Assessments of Projected Appropriate Monetary Policy, March 2019." https://www.federalreserve.gov/monetarypolicy/files/fomcprojtabl20190320.pdf.

Gustavsson, Magnus J., and Pär Österholm. 2007. "Does Unemployment Hysteresis Equal Employment Hysteresis?" *Economic Record* 83, no. 261: 159–73.

Hagedorn, Marcus, and Iourii Manovskii. 2013. "Job Selection and Wages over the Business Cycle." *American Economic Review* 103, no. 2: 771–803.

Hall, Robert E. 1991. "Labor Demand, Labor Supply, and Employment Volatility," In *NBER Macroeconomic Annual 1991*, edited by Olivier Jean Blanchard and Stanley Fischer. Cambridge, Mass.: MIT Press.

Holzer, Harry J., Steven Raphael, and Michael A. Stoll. 2006 "Employers in the Boom: How Did the Hiring of Less-Skilled Workers Change during the 1990s?" *Review of Economics and Statistics* 88, no. 2: 283–99.

Hotchkiss, Julie L., and Robert E. Moore. 2018. "Some Like It Hot: Assessing Longer-Term Labor Market Benefits from a High-Pressure Economy." Working Paper 2018-1b. Atlanta: Federal Reserve Bank of Atlanta.

Hoynes, Hilary. 2000. “The Employment and Earnings of Less Skilled Workers over the Business Cycle.” In *Finding Jobs: Work and Welfare Reform*, edited by Rebecca Blank and David Card. New York: Russell Sage Foundation.

Jefferson, Philip N. 2008. “Educational Attainment and the Cyclical Sensitivity of Employment.” *Journal of Business and Economic Statistics* 26, no. 4: 526–35.

Kahn, Lisa B. 2010 “The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy.” *Labour Economics* 17, no. 2: 303–16.

Katz, Lawrence, and Alan Krueger. 1999. “The High-Pressure U.S. Labor Market of the 1990s.” *Brookings Papers on Economic Activity*, no. 1, 1–87. https://www.brookings.edu/wp-content/uploads/1999/01/1999a_bpea_katz.pdf.

Kiley, Michael T. 2015. “An Evaluation of the Inflationary Pressure Associated with Short- and Long-Term Unemployment.” *Economics Letters* 137: 5–9.

Kondo, Ayako. 2015. “Differential Effects of Graduating during a Recession across Gender and Race.” *IZA Journal of Labor Economics* 4, no. 23. https://izajole.springeropen.com/track/pdf/10.1186/s40172-015-0040-6.

Leduc, Sylvain, Chitra Marti, and Daniel Wilson. 2019. “Does Ultra-Low Unemployment Spur Rapid Wage Growth?” *FRBSF Economic Letter* (Federal Reserve Bank of San Francicso) 2019-02, January 14.

Leduc, Sylvain, and Daniel J. Wilson. 2017. “Has the Wage Phillips Curve Gone Dormant?” *FRBSF Economic Letter* 2017-30, October 16.

Modestino, Alicia Sasser, Daniel Shoag, and Joshua Balance. 2016. “Downskilling: Changes in Employer Skill Requirements over the Business Cycle.” *Labour Economics* 41: 333–47.

Okun, Arthur M. 1973. “Upward Mobility in a High-Pressure Economy.” *Brookings Papers on Economic Activity*, no. 1, 207–52. https://www.brookings.edu/wp-content/uploads/1973/01/1973a_bpea_okun_fellner_greenspan.pdf.

Oreopoulos, Philip, Till von Wachter, and Andrew Heisz. 2012. “The Short- and Long-Term Career Effects of Graduating in a Recession.” *American Economic Journal: Applied Economics* 4, no. 1: 1–29.

Orphanides, Athanasios, and John C. Williams. 2013. “Monetary Policy Mistakes and the Evolution of Inflation Expectations.” In *The Great Inflation: The Rebirth of Modern Central Banking*, edited by Michael D. Bordo and Athanasios Orphanides. Chicago: University of Chicago Press.

Parker, Jonathan A., and Annette Vissing-Jorgensen. 2010. “The Increase in Income Cyclicality of High-Income Households and Its Relation to the Rise in Top Income Shares.” *Brookings Papers on Economic Activity*, Fall, 1–70. https://www.brookings.edu/wp-content/uploads/2010/09/2010b_bpea_parker.pdf.

Reifschneider, Dave, William Wascher, and David Wilcox. 2015. “Aggregate Supply in the United States: Recent Developments and Implications for the Conduct of Monetary Policy.” *IMF Economic Review* 63, no. 1: 71–109.

Romer, Christina D., and David Romer. 1999. “Monetary Policy and the Well-Being of the Poor.” *Economic Review–Federal Reserve Bank of Kansas City,* First Quarter, 21–49.

Smith, Christopher L. 2014. “The Effect of Labor Slack on Wages: Evidence from State-level Relationships.” *FEDS Notes*, Federal Reserve Board of Governors, June 2. https://www.federalreserve.gov/econresdata/notes/feds-notes/2014/effect-of-labor-slack-on-wages-evidence-from-state-level-relationships-20140602.html.

Song, Frank M., and Yangru Wu. 1997. “Hysteresis in Unemployment: Evidence from 48 U.S. States.” *Economic Inquiry* 35, no. 2: 235–43.

Staiger, Douglas, Mark Watson, and James Stock. 2001. “Prices, Wages and the U.S. NAIRU in the 1990s.” In *The Roaring Nineties*, edited by Alan Krueger and Robert Solow. New York: Russell Sage Foundation.

Weingarden, Alison. 2017. “Labor Market Outcomes in Metropolitan and Non-Metropolitan Areas: Signs of Growing Disparities.” *FEDS Notes*, Federal Reserve Board of Governors, September 25. https://www.federalreserve.gov/econres/notes/feds-notes/labor-market-outcomes-in-metropolitan-and-non-metropolitan-areas-signs-of-growing-disparities-20170925.htm.

Wilson, Valerie. 2015. “The Impact of Full Employment on African American Employment and Wages,” Economic Policy Institute. https://www.epi.org/publication/the-impact-of-full-employment-on-african-american-employment-and-wages/.

Yagan, Danny. Forthcoming. “Employment Hysteresis from the Great Recession.” *Journal of Political Economy*.

Comments and Discussion

COMMENT BY

JULIE L. HOTCHKISS Worse labor market outcomes among racial and ethnic minorities, the less educated, and to a certain extent, females, seem to have become an unalterable fact of the U.S. labor market. For decades, these "disadvantaged" workers have been getting the short end of the stick with respect to unemployment rates, labor force participation, wages, and hours of work. Economists have long sought explanations and solutions for the significant gaps in labor market outcomes between the disadvantaged and advantaged—or, rather, whites, the educated, and males. The identification of large unemployment disparities as a social issue has a long history, dating back at least to George Perry's (1970) identification of structural factors playing a role in the relationship between what level of unemployment can be attained at a given level of inflation, and Robert Hall's (1970) consideration of whether the notion of "normal" unemployment differs by race and gender. This paper by Stephanie Aaronson, Mary Daly, William Wascher, and David Wilcox follows in this tradition with a very specific goal: is there any evidence that periods of particularly strong labor markets can put a dent in these persistent gaps?

For the authors, all of whom have or have had policymaking or advising positions within the Federal Reserve, this question is not merely academic. With full employment being a legislated goal of the Federal Reserve, critics have argued that monetary policymakers should consider a more inclusive definition of full employment that places significant weight on the labor market outcomes of the disadvantaged. One way to do this is to adopt policies that encourage and prolong a "hot" or "high-pressure" economy. A hot economy, defined as one in which the unemployment rate falls below the Congressional Budget Office's (CBO) estimated sustainable unemployment rate, runs the risk (among other things, as identified by the authors) of increasing inflation. So the question is not only whether there is evidence that a hot economy can help to close the gap in labor market outcomes, but also whether the degree of success expected is worth the risks it entails. The prevailing consensus on this point, consistent with results presented by Aaronson and her colleagues, is that though there is some evidence that a hot economy disproportionately improves the contemporaneous labor market outcomes of disadvantaged workers, the benefit neither sticks nor is it able to undo the disproportionate harm disadvantaged workers suffer during a "cold" economy (for example, see Hotchkiss and Moore 2018; and Fallick and Krolikowski 2018).

The authors provide evidence that is consistent with the literature they cite. In these remarks, I first take their results one step further to illustrate why a hot economy, alone, has not been effective in closing labor market gaps between advantaged and disadvantaged workers. Then—using their methodology, but with a different data set that allows observing individuals over many years and multiple business cycles—I offer additional evidence that the positive impact of hot economies does not reach very far into the future. And finally, I return to Okun's own words to interpret today's empirical evidence for policy considerations.

A LACK OF PROGRESS ACROSS BUSINESS CYCLES In drawing conclusions from the results presented by Aaronson and her colleagues, one needs to keep in mind that cold economic periods typically last longer and are more intense than hot periods. This can easily be seen in figure 1 of their paper. Even if the marginal impact of a negative unemployment gap (a hot economic period) exceeds the marginal impact of a positive unemployment gap for a particular disadvantaged group, the net total impact over the business cycle is not likely to benefit the disadvantaged group. As a concrete example of this, I use the estimation results from the authors' table 3. Their table 3 presents the average differential impact on the unemployment rate for black males and females, relative to white males and females, from an increase of 1 percentage point in the unemployment rate gap during four business cycle episodes that include both cold and hot periods. Their estimation also allows for a differential impact of a gap during hot periods. My table 1 summarizes each business cycle's cold and hot periods and the average and total differential impacts for blacks, relative to whites, from exposure to these hot and cold periods. Note that the business cycle starting in 2008 is extended through 2022 as projected by the authors in their figure 1.

The first thing to notice in my table 1 is that in each business cycle, the number of cold quarters exceeds the number of hot quarters, although the 1990–2001 business cycle came close to being an exception, with 23 cold

Table 1. Summary of Cold and Hot Periods by Business Cycle and the Differential Impact on the Unemployment Rate for Black Males and Females, Relative to White Males and Females, from a Rising Unemployment Rate Gap[a]

Business cycle	*Business cycle details*	*Differential impact on unemployment rate, blacks versus whites*	
		Males	*Females*
1980:Q1–1990:Q3			
Number of cold quarters	32		
Number of hot quarters	11		
Average gap—cold period	1.668		
Average gap—hot period	–0.409		
Average differential—cold		1.425	0.926
Average differential—hot		–0.175	0.029
Total differential—cold		45.588	29.631
Total differential—hot		–1.927	0.320
1990:Q4–2001:Q3			
Number of cold quarters	23		
Number of hot quarters	21		
Average gap—cold period	0.974		
Average gap—hot period	–0.768		
Average differential—cold		0.840	0.641
Average differential—hot		–0.810	–0.643
Total differential—cold		19.315	14.744
Total differential—hot		–17.015	–13.499
2001:Q4–2007:Q4			
Number of cold quarters	16		
Number of hot quarters	9		
Average gap—cold period	0.552		
Average gap—hot period	–0.315		
Average differential—cold		0.140	0.185
Average differential—hot		–0.241	–0.866
Total differential—cold		2.242	2.957
Total differential—hot		–2.168	–7.794
2008:Q1–2022:Q4			
Number of cold quarters	37		
Number of hot quarters	23		
Average gap—cold period	2.568		
Average gap—hot period	–0.676		
Average differential—cold		2.324	1.138
Average differential—hot		–1.220	–1.066
Total differential—cold		86.000	42.097
Total differential—hot		–28.055	–24.509

Sources: Bureau of Labor Statistics; Congressional Budget Office; author's calculations.

a. Average difference for blacks during hot and cold periods within each business cycle is calculated as follows: $\overline{Diff\,(gap)}_{Cold} = \frac{1}{N_C}\sum_{t=1}^{N_C}[\widehat{\alpha_1} * gap_t]$ and $\overline{Diff\,(gap)}_{Hot} = \frac{1}{N_H}\sum_{t=1}^{N_H}[\widehat{\alpha_1} * gap_t + \widehat{\alpha_2} * gap_t]$, where $\widehat{\alpha_1}$ and $\widehat{\alpha_2}$ are taken from table 3 in the paper by Aaronson and her colleagues, N_C is the number of "cold" quarters in the business cycle, N_H is the number of "hot" quarters in the business cycle, and gap_t is calculated using the difference between the aggregate unemployment rate from the BLS and the natural rate of unemployment from the Congressional Budget Office (obtained from https://fred.stlouisfed.org/series/NROU) in quarter t. Estimates through 2022 use estimated parameters from the fourth business cycle. The total impact of each business cycle's cold (or hot) period is calculated simply as the average difference during the cold (or hot) period times the number of cold (or hot) quarters.

quarters and 21 hot quarters. The second thing to notice is that, for all four business cycles, the average unemployment rate gap during cold quarters is greater than the absolute value of the average gap during hot quarters. For example, during the extended 2008 business cycle, the average cold period gap is 2.57 percentage points and the average hot period gap is -0.68 percentage point. In other words, cold periods are more intense labor market environments, on average, than hot periods.

Moving to the estimated differential effects for blacks versus whites, the third thing to notice in my table 1 is that, even though blacks tend to have a "higher-beta" experience for these outcomes during hot periods than during cold periods (as noted by Aaronson and her colleagues), the greater intensity of cold periods produces an average differential impact of the gap that is greater during cold periods than during hot periods. It is the combined influence of their longer length and greater average differential impact on the labor market experiences of blacks relative to whites that results in cold periods having larger total differential effects on blacks than hot periods.[1] This is shown in the last two rows for each business cycle in my table 1.

Purely as a thought experiment, we can use the information in my table 1 to estimate how many additional quarters a hot period would have to last (at the same average hot period differential impact) or what additional hot period differential impact it would take (at the same number of hot period quarters) for the total hot and cold differential effects for blacks versus whites to be equal across the business cycle (this would mean just breaking even for blacks). Using the unemployment rate for males in the current business cycle as an example, this exercise indicates that even if the current hot period extends through 2022, it would need to continue for another 48 quarters (12 years) beyond that, or the average differential hot period impact would need to have been an additional 2.5 percentage points lower unemployment rate for blacks, relative to whites, during the hot period through 2022 in order to wipe out the negative impact of the business cycle's cold period.[2] A less volatile business cycle, such as 2001–7, would have only required less than one-quarter of additional high-pressure

1. The exception to this in my table 1 is for women across the 2001 business cycle, where the absolute value of the total differential impact on the unemployment rate during the hot period (–7.79) exceeds the differential impact during the cold period (2.96).

2. To equalize the experience for black and white females over this extended business cycle, the hot period would have to last about four years beyond 2022 or the average differential hot period impact for black females would have to be 0.75 percentage point lower unemployment rate.

exposure or an additional 0.01 percentage point differential impact of the gap on the unemployment rate for black males.[3]

The implication of Aaronson and her colleagues' results across the full business cycle is that disadvantaged workers cannot seem to get a leg up. Stronger gains during hot economic periods are typically wiped out by even stronger setbacks during cold periods. This is consistent with my own research, using different data and methodology.

A QUESTION OF PERSISTENCE In light of the net negative impact of business cycles for disadvantaged groups, it would be hard to believe that hot economic environments have a cumulative, positive long-term impact on reducing labor market outcome gaps. In an effort to determine whether there is longer-term improvement in unemployment rates and nonparticipation among disadvantaged demographic groups, the authors present evidence (through tests for a unit root) for a trend in nonparticipation for all groups, but only in a few cases in the unemployment rate. As they point out, however, nothing in their analysis ties the presence of a unit root to exposure to a hot economic environment. In their excellent review of the literature, the authors point to various analyses using panel data to find evidence of negative hysteresis, or persistence, from cold economic events, such as entering the labor force for the first time during a recession.

In my own research, I have found little evidence of lasting positive effects from exposure to a hot economy. Complementing that earlier research, I loosely apply the authors' methodology to investigate this question of persistence using panel data, which tracks individuals over many years and multiple economic environments.[4] The analysis here also differs from that of Aaronson and her colleagues, in that it uses state-specific, long-term unemployment rates to calculate the unemployment gap—that is, the deviation of the state's unemployment rate from its long-term unemployment

3. Extra hot period number of quarters for a given business cycle needed to overcome the total impact of the cold period is calculated as follows: $N_H^{extra} = \frac{-\sum_{t=1}^{N_C} Diff_t^C}{\overline{Diff}_H} - N_H$, where $Diff_t^C$ is the estimated differential unemployment rate predicted for blacks from the unemployment gap in quarter t, $\overline{Diff}_H$ is the average differential unemployment rate experienced by blacks relative to whites during the hot period of the business cycle, and N_H is the actual number of quarters the hot period of the business cycle lasted. The formula for calculating the extra differential impact needed is given by $\overline{Diff}_H^{extra} = \frac{-\sum_{t=1}^{N_C} Diff_t^C}{N_H} - \overline{Diff}_H$.

4. The 1979 and 1997 cohorts of the National Longitudinal Surveys of Youth, covering the years 1982–2014, are used. The analysis is restricted to age 25–57 years (the maximum age in the sample).

rate, based the CBO's national long-term unemployment rate and deviations of the state's long-run average from the national long-run average; details can be found in my 2018 paper with Robert Moore.[5] As in the paper by Aaronson and her colleagues, the gap can either be positive (indicating a cold economy) or negative (indicating a hot economy), and the impact of the gap is allowed to differ across cold and hot environments.

Here, I consider two labor market outcomes: unemployment and real hourly wages. Each labor market outcome, $LMoutcome_{i,s,t}$, of person i in state s in year t is expressed as a function of the person's individual demographics (age, race, education, and gender, which all enter as group dummies) and the current and lagged values of the unemployment gap ($GAP_{i,s,t-j}$). The unemployment gap enters separately, interacts with each demographic characteristic, and is allowed to affect outcomes differently during hot economic environments through $HotDum_{t-j}$:

$$
\begin{aligned}
(1)\quad & LMoutcome_{i,s,t} \\
&= \alpha + \sum_{k=2}^{4}\left\{AGE_i^k\left(\delta_{1k} + \sum_{j=0,2,4} GAP_{i,s,t-j}\left[\delta_{2k,j} + \delta_{3k,j} HotDum_{t-j}\right]\right)\right\} \\
&\quad + \sum_{k=2}^{3}\left\{RACE_i^k\left(\beta_{1k} + \sum_{j=0,2,4} GAP_{i,s,t-j}\left[\beta_{2k,j} + \beta_{3k,j} HotDum_{t-j}\right]\right)\right\} \\
&\quad + \sum_{k=2}^{3}\left\{EDUC_i^k\left(\varphi_{1k} + \sum_{j=0,2,4} GAP_{i,s,t-j}\left[\varphi_{2k,j} + \varphi_{3k,j} HotDum_{t-j}\right]\right)\right\} \\
&\quad + MALE_i\left\{\theta_1 + \sum_{j=0,2,4} GAP_{i,s,t-j}\left[\theta_{2j} + \theta_{3j} HotDum_{t-j}\right]\right\} \\
&\quad + \sum_{j=0,2,4} GAP_{i,s,t-j}\left[\rho_{1j} + \rho_{2j} HotDum_{t-j}\right] + \tau_t + \sigma_s + \pi_i + \varepsilon_{i,s,t}.^{6}
\end{aligned}
$$

Marginal effects for the impact of a change in the gap (contemporaneous and lagged) on the two labor market outcomes are reported in my table 2, for the full sample and by race.[7] First, note that the "higher-beta" experience for blacks is most evident with the lagged gaps. For example,

5. The pattern of results using a comparison of state unemployment rates to the national long-term unemployment rate is similar to those presented here.

6. Although this estimation includes individual fixed effects, robustness checks indicate that this does not make much difference to the point estimates. State and year dummies are included, and errors are clustered at the state level. Analysis is restricted to individuals with minimal labor market attachment. Lags of two years are considered because later National Longitudinal Surveys of Youth surveys are done every two years.

7. Weekly hours and labor force participation were also explored as additional labor market outcomes and produce a similar pattern, but less precise estimates. The pattern of results in my table 2 are also generally consistent across advantaged and disadvantaged age and education groups.

Table 2. Marginal Effects for the Impact of Gap Changes on Unemployment and the Log of Real Hourly Wages[a]

	Current gap		*Gap lagged 2 years*		*Gap lagged 4 years*	
Outcome/demographic group	*Positive gap (cold economy)*	*Negative gap (hot economy)*	*Positive gap (cold economy)*	*Negative gap (hot economy)*	*Positive gap (cold economy)*	*Negative gap (hot economy)*
Impact on unemployment						
Full sample	0.0023**	0.0004	0.0072***	0.0044***	0.0031**	0.0007
	[0.0008]	[0.0022]	[0.0008]	[0.0009]	[0.0010]	[0.0016]
White, non-Hispanic	0.0016	0.0033	0.0068***	0.0041**	0.0020*	–0.0003
	[0.0009]	[0.0021]	[0.0010]	[0.0013]	[0.0009]	[0.0018]
Hispanic	0.0024	–0.0011	0.0055***	0.0035	0.0034*	0.0055
	[0.0013]	[0.0040]	[0.0013]	[0.0023]	[0.0017]	[0.0042]
Black, non-Hispanic	0.0035***	–0.0042	0.0090***	0.0054*	0.0050***	–0.0003
	[0.0010]	[0.0033]	[0.0011]	[0.0021]	[0.0015]	[0.0020]
Impact on log of real hourly earnings						
Full sample	0.0041	–0.0073	–0.0035	–0.0154***	–0.0111***	–0.0012
	[0.0029]	[0.0052]	[0.0027]	[0.0031]	[0.0020]	[0.0051]
White, non-Hispanic	0.0055	–0.0039	–0.0034	–0.0103*	–0.0104***	–0.0031
	[0.0030]	[0.0069]	[0.0028]	[0.0042]	[0.0019]	[0.0054]
Hispanic	0.0048	–0.0219***	–0.001	–0.0190*	–0.0109***	–0.011
	[0.0037]	[0.0061]	[0.0034]	[0.0079]	[0.0027]	[0.0064]
Black, non-Hispanic	0.0011	–0.0038	–0.0053	–0.0226***	–0.0125***	0.009
	[0.0028]	[0.0054]	[0.0030]	[0.0039]	[0.0029]	[0.0075]

Sources: National Longitudinal Survey of Youth; author's calculations.

a. Data are the 1979 and 1997 cohorts of the National Longitudinal Surveys of Youth covering the years 1982–2014. Unemployment is measured as the share of time during the year in the labor market spent unemployed. Real hourly pay is in 2014 dollars.

a decrease of 1 percentage point in the gap during a hot economy two years earlier decreases the share of time in the labor force during the year that blacks spend being unemployed, on average, by 0.5 percentage point, whereas the average time spent being unemployed decreases by 0.4 percentage point for whites. There is also a larger impact on average hourly earnings for blacks.

A larger contemporaneous gap increases unemployment (during a cold economy) but does not significantly affect hourly earnings. A larger gap two years earlier increases current unemployment experience (in both a hot and cold economy) and decreases current real hourly earnings (only in a hot economy). In other words, a hotter economy (meaning a more negative gap) two years earlier will have a positive impact on current hourly wages. Higher unemployment gaps four years earlier increase the current unemployment experience and decrease current real hourly earnings, in both cases only in a cold economy. Gaps longer than four years earlier were generally not found to be statistically significant. The conclusion is that exposure to a hot economic environment does not appear to have a particularly long-lasting impact on individual labor market outcomes.

IMPLICATIONS FOR POLICYMAKING Arthur Okun's preoccupation with the relationship between the labor market and output is predicated on his assumption "that idle labor is a satisfactory measure of all idle resources" (Okun 1962, 6). The channel through which a hot labor market translates into higher economic output is through increased individual productivity—what Okun (1973) referred to as "cyclical upgrading." This upgrading of productivity can take place, according to Okun, in three ways. First, employment in more productive industries is more volatile across the business cycle; during a hot economy, workers shift from less productive to more productive industries because that is where the greater demand for labor is concentrated. Second, workers experience upward movement by climbing productivity ladders within firms. And third, geographic mobility will allow workers to move from lower-income to higher-income regions during periods of high demand. Okun (1973, 227) only provides empirical evidence for the first of these three potential channels, but he speculates that "the skills accumulated during years of employment in [higher-productivity jobs made possible during hot economies] may make workers much more adaptable for good jobs elsewhere." The implication is that a hot economy provides an opportunity (with emphasis on "opportunity") for the effects of cyclical upgrading to be long-lasting.

Aaronson and her colleagues provide and cite significant evidence that hot economies have a disproportionately positive, contemporaneous impact on labor market outcomes among disadvantaged workers. However, the overwhelming evidence in the literature so far is that exposure to a hot economy does not have a lasting positive impact on individual labor market outcomes—including unemployment, labor force participation, hours of work, and, Okun's favorite, wages. The general comments offered on Okun's (1973) paper were by and large quite skeptical about the permanence of the cyclical upgrading laid out by Okun. The general discussion notes that Okun responded that "he was claiming not that all the upgrading effects he uncovered were permanent, but only that they lasted long enough to be *important*" (quoted by Okun 1973, 259; emphasis added). This point is crucial for policy considerations.

It is clear that in his paper, Okun did not suggest that a "high-pressure policy" will, by itself, permanently reduce labor market gaps between advantaged and disadvantaged workers. Just creating more good jobs is not enough. He calls for "manpower programs" (that is, policymakers) to take advantage of hot economic environments to "incorporate a major effort to instill training and the basis for upgrading [skills], rather than merely create more [good] jobs" (Okun 1973, 245). He goes on to say, "Barriers to entry into good jobs may be swept away most easily when market forces are making racial and sexual discrimination costly to employers" (Okun 1973, 245).

Given that the results presented by Aaronson and her colleagues indicate that there is also no relative improvement in labor market outcomes (even from contemporaneous exposure) across the business cycle for disadvantaged workers (see my table 1), policymakers are clearly not taking advantage of hot economic environments to break down barriers to "good jobs." In fact, the contemporaneous hot economy bonus accruing to disadvantaged workers, which will quickly disappear during the next economic downturn, indicates that the bonus is more a suspension of prejudice and discriminatory behavior than an upgrading of individual productivity. The implication for monetary policy is that without a coordinated effort from makers of social policy to capitalize on employers' desperate need for labor to break down forces of discrimination and prejudice, accommodation of a high-pressure economy for the purposes of long-term improvement in labor market gaps will be ineffective.

REFERENCES FOR THE HOTCHKISS COMMENT

Fallick, Bruce, and Pawel Krolikowski. 2018. "Hysteresis in Employment among Disadvantaged Workers." Working Paper WP 18-01. Cleveland: Federal Reserve Bank of Cleveland. https://www.clevelandfed.org/newsroom%20 and%20events/publications/working%20papers/2018%20working%20papers/wp%201801%20hysteresis%20in%20employment%20among%20disadvantaged%20workers.

Hall, Robert E. 1970. "Why Is the Unemployment Rate So High at Full Employment?" *Brookings Papers on Economic Activity* 1, no. 3: 369–402.

Hotchkiss, Julie L., and Robert E. Moore. 2018. "Some Like It Hot: Assessing Longer-Term Labor Market Benefits from a High-Pressure Economy." Working Paper 2018-1. Atlanta: Federal Reserve Bank of Atlanta. https://www.frbatlanta.org:443/research/publications/wp/2018/01b-assessing-longer-term-labor-market-benefits-from-a-high-pressure-economy-2018-10-23.

Okun, Arthur M. 1962. "Potential GNP: Its Measurement and Significance." *American Statistical Association: Proceedings of the Business and Economic Statistics Section*, 98–104.

———. 1973. "Upward Mobility in a High-Pressure Economy." *Brookings Papers on Economic Activity* 1973, no. 1: 207–61. https://doi.org/10.2307/2534087.

Perry, George L. 1970. "Changing Labor Markets and Inflation." *Brookings Papers on Economic Activity* 1, no. 3: 411–41.

COMMENT BY

JUSTIN WOLFERS This paper by Stephanie Aaronson, Mary Daly, William Wascher, and David Wilcox explores how the business cycle shapes the labor market outcomes of different demographic groups. It updates and expands upon an existing literature, which has typically found that recessions do more harm to the labor market prospects of disadvantaged groups than to those of others, and that economic expansions also do more to boost their labor market prospects. That is, the labor market outcomes of disadvantaged groups tend to be especially procyclical. A finance economist would say that these are "high-beta" groups.

This new paper makes two contributions to this literature, and my comment responds to each. First, Aaronson and her colleagues update and confirm earlier findings that the labor market prospects of certain disadvantaged groups are especially sensitive to business cycle conditions. The first part of my comment explores these findings further, showing that they reveal a particularly interesting structure. Second, the authors look for evidence of an asymmetry, asking whether the boost these groups get from a "hot" labor market is larger than the harm done by an equally "cold" labor market. Although the authors strike a mildly optimistic tone, arguing that they uncovered some suggestive evidence in favor of their hypothesis, I am less optimistic. That is because the second part of my comment expands the authors' analysis beyond the United States, finding no systematic evidence in favor of their hypothesis. An additional strength of this paper is that it expands the array of labor market measures that are typically analyzed, evaluating the cyclicality of not just unemployment but also participation rates, average hourly earnings, own earnings, and household earnings. For the sake of brevity, I focus my comment only on the unemployment rate.

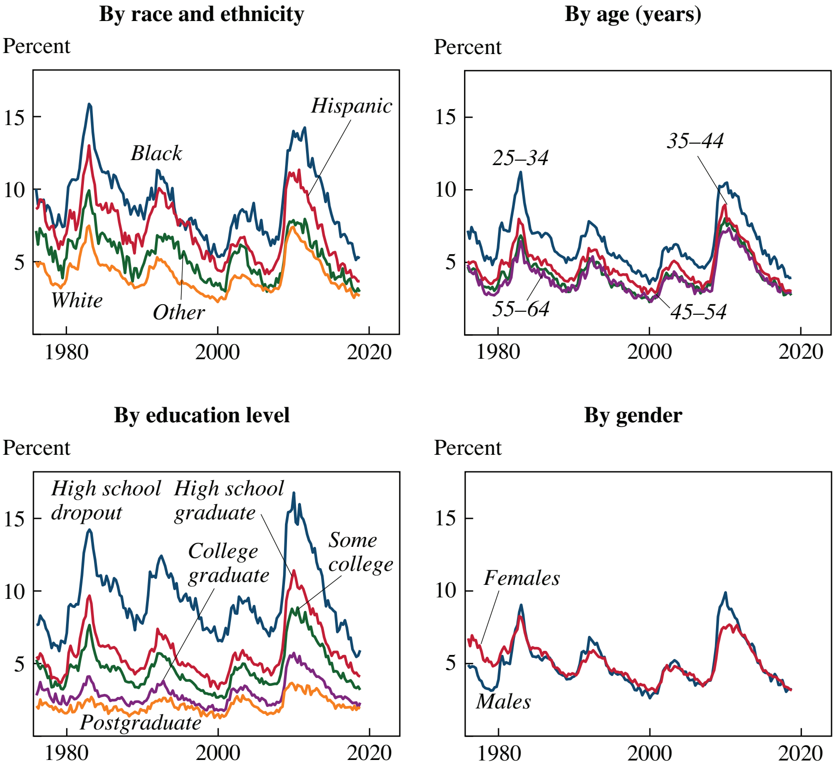
BENEFITS OF A HIGH-PRESSURE LABOR MARKET My figure 1 illustrates the key ideas in the paper by Aaronson and her colleagues, graphing the unemployment rate by race and ethnicity, by gender, by education, and by age. The data I used to construct these plots—and throughout this comment—are largely the same as those used by the authors. Following their approach, I start with micro data on labor market outcomes from the Current Population Survey (CPS), focusing on those age 25–64 years over the period 1976–2018.[1] I also expand a little on their analysis. Though the authors only show results for three race/ethnicity groups (non-Hispanic blacks, non-Hispanic whites, and Hispanics), I also include the "other" category. And though they show only a coarse categorization of education into three groups, I separate out high school graduates from high school dropouts, and I also separate out those with only a college degree from those who have studied for postgraduate degrees. In addition, though the authors only analyze differences by gender, race/ethnicity, and education, I also separate out four age groups, analyzing those age 25–34, 45–44, 45–54, and 55–64 years.

The top left panel of my figure 1 shows that the unemployment rate for blacks is clearly both higher and more cyclically sensitive than it is for whites. Between these lies the unemployment rate for Hispanics, which is both lower and less cyclical than that of blacks, and higher and more cyclical than that of whites. (The CPS also contains a residual "other" group, whose unemployment rate is a bit lower and less cyclical than that of Hispanics but, again, higher and more cyclical than that of whites.)

The next panel of my figure 1 shows differences by age, a dimension that Aaronson and her colleagues do not analyze in much detail. The unemployment rates for those age 45–54 and 55–64 are virtually identical. The unemployment rate for those age 35–44 is slightly higher, and slightly more cyclical. The more notable difference arises with the youngest age

1. I drew these micro data from the cleaned and harmonized data file maintained by the Federal Reserve Bank of Kansas City, and I then aggregated them into a quarterly series, which I seasonally adjusted using a simple ratio–to–moving average filter.

Figure 1. Unemployment Rates by Demographic Group, 1976–2017



Source: Author's calculations, using the Current Population Survey.

group—those age 25–34—whose unemployment rate is both substantially higher and substantially more cyclical than that of any other age group.

The starkest differences are across education levels, which are shown in the figure's lower left panel. The unemployment rate for high school dropouts is both higher and more cyclical than that of high school graduates, which in turn is higher and more cyclical than that of those with some college, which in turn is higher and more cyclical than that of college graduates. Indeed, it is quite striking just how low and steady the unemployment rate of those with postgraduate degrees is, given that throughout this entire period, it rose or fell within a range that was only 1½ percentage points above or below its mean level.

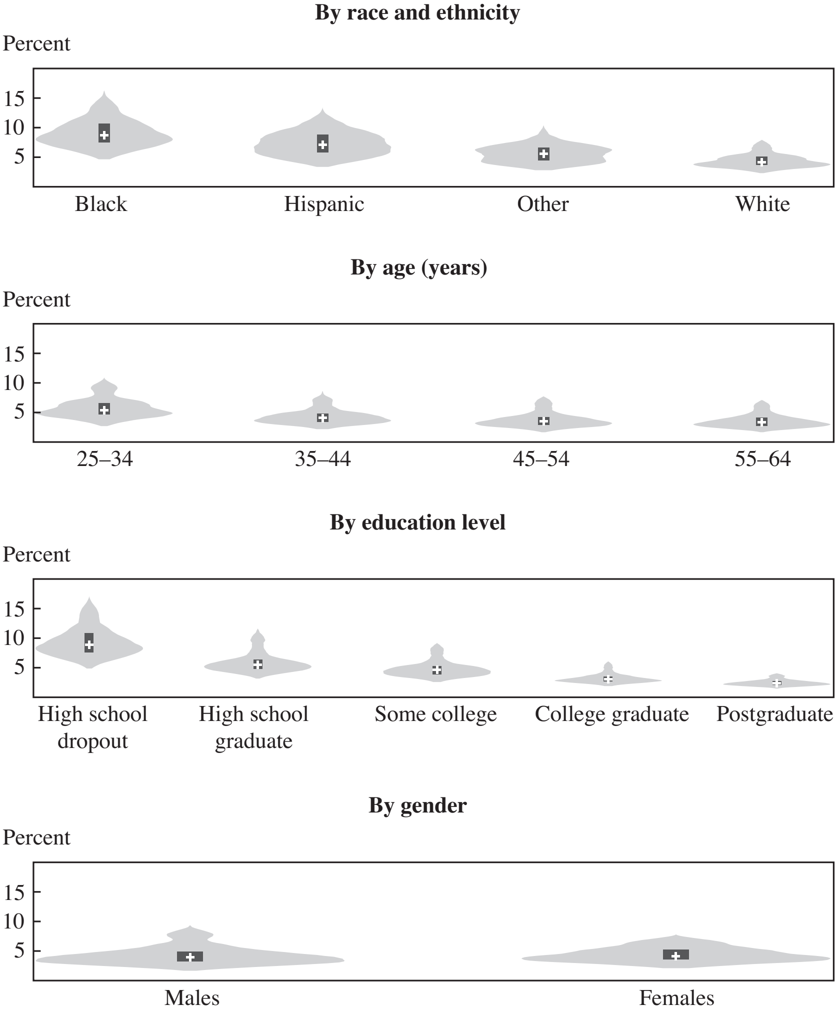
The common theme here is that disadvantaged groups—racial and ethnic minorities, young people, and those with less education—all appear to have both higher and more cyclically sensitive unemployment rates. Gender differences in both the level and cyclicality of unemployment rates are much smaller.

My figure 2 shows a closely related idea, in a different format, presenting violin plots for each demographic group. As a reminder, the light gray bell curves (the violins) show kernel density estimates of the probability density function of the unemployment rate; the black bar in the middle illustrates the interquartile range of unemployment for that group; and the white cross illustrates the median unemployment rate for that group. In each panel, the plots shown on the left summarize the distribution of unemployment rates for disadvantaged groups—blacks and Hispanics, younger people, and those with less education. In each case, these groups have both higher unemployment rates on average and also more variable unemployment rates, reflecting the greater amplitude of their cyclical variation. As you look across to the right within each panel, the plots for more advantaged groups—whites, older people, and those with college degrees—reveal unemployment rates that are both lower on average and much less variable.

Taken together, this evidence suggests that there is a demographic dividend from running a high-pressure labor market—not only do all groups benefit, but those who have historically been disadvantaged benefit the most. In this sense, the findings confirm Arthur Okun's (1973, 246) argument that "the greater diffusion of opportunity and of upward mobility in a full-utilization economy is a vital social benefit; and that benefit helps explain why the pursuit of full employment is an integral part of a liberal's creed."

LINKING THE LEVEL AND CYCLICALITY OF UNEMPLOYMENT Implicit in Okun's idea is the notion that somehow the factors that make some people more susceptible to unemployment, on average, also make them more susceptible to cyclical fluctuations. But the evidence the literature has accumulated on this point so far is largely informal and qualitative, essentially just pairing the observation that those groups with more cyclical unemployment rates—blacks, Hispanics, younger workers, and those with less education—have also tended to have higher unemployment rates. What has not previously been tested is the more precise quantitative prediction: If disadvantage explains both the susceptibility of a person or group to unemployment, and their susceptibility to cyclical fluctuations, then from a statistical perspective, a single index can explain both phenomena. This single index property suggests that the cyclicality of a demographic group's unemployment should rise in proportion with its unemployment rate. In what follows, I extend the research of Aaronson and her colleagues to examine whether it is consistent with this additional quantitative prediction.

Figure 2. The Distribution of Unemployment Rates, 1976–2017[a]



Source: Author's calculations, using the Current Population Survey.

a. The probability density function is shown in light gray; the 25th–75th percentiles are highlighted in black; and the median is shown as a white cross.

A micro data perspective. I compiled the CPS micro data for the United States for the period and sample of the authors' study, and ran this regression:

$$U_{i,t} = \begin{pmatrix} \sum_r \alpha_r I(race_i = r) + \sum_a \alpha_a I(age_i = a) \\ + \sum_e \alpha_e I(educ_i = e) + \sum_s \alpha_s I(sex_i = s) \end{pmatrix} + \begin{bmatrix} \sum_r \beta_r I(race_i = r) + \sum_a \beta_a I(age_i = a) \\ + \sum_e \beta_e I(educ_i = e) + \sum_s \beta_s I(sex_i = s) \end{bmatrix} \times \underbrace{(U_t - U_t^*)}_{\substack{\text{Aggregate} \\ \text{unemployment gap}}},$$

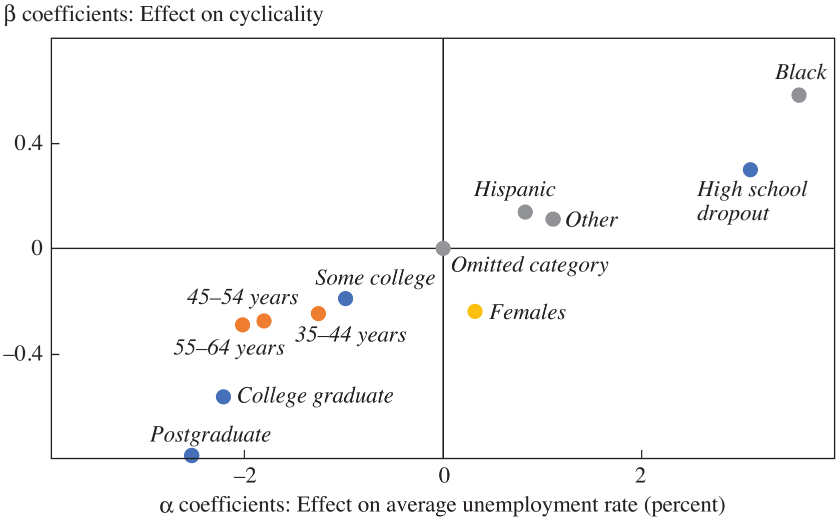
where the dependent variable, $U_{i,t}$, is a binary variable set to 1 if individual i is unemployed at time t, and 0 if he or she is employed (those who are not in the labor force are excluded from the sample). The α coefficients effectively describe how the average risk of unemployment varies, depending on each person's demographic characteristics. In particular, it describes differences in the unemployment rate that might occur when the aggregate unemployment rate is equal to the equilibrium rate. By contrast, the βs describe how sensitive the unemployment risk of people with different demographic characteristics is to the state of the business cycle, which, following Aaronson and her colleagues, I measure as the gap between the national unemployment rate, U_t, and the equilibrium unemployment rate calculated by the Congressional Budget Office, U_t^*.

The point is that the αs describe how demographic characteristics shape the average level of unemployment, while the βs describe its cyclicality. Importantly, the idea that both are determined by a common factor called "disadvantage" suggests that any characteristic c (which might refer to race, age, education, or gender) that leads to a higher α_c will also lead to a commensurately higher β_c. And so, rather than presenting these regression results in a standard table, my figure 3 graphs them, showing the β_c for each characteristic against the corresponding α_c.

The findings clearly are consistent with the idea that demographic characteristics that lead to higher unemployment also lead to more cyclical unemployment. This pattern can be seen both within each demographic characteristic, and also between them. Indeed, across these characteristics, the correlation between the estimates of α_c and β_c is 0.93 (and with a t statistic of 7.9, this meets standard metrics for statistical significance).

A macro data perspective. An alternative approach to the same question considers the unemployment rates of quite specific demographic groups.

Figure 3. Regression Results Showing How Demographic Characteristics Shape the Average Level and Cyclicality of Unemployment



Source: Author's calculations, using the Current Population Survey.

For this, I used the same CPS micro data to construct seasonally adjusted estimates of the quarterly unemployment rate for quite fine partitions of the workforce by race and ethnicity, gender, education, and age. This led to a total of 160 separate unemployment rates (4 race/ethnicity groups × 2 genders × 5 education groups × 4 age groups). The advantage of this approach is that it accounts for all possible interactions between these demographic characteristics.

For each of these narrowly defined demographic groups, g, I ran a simple regression of the form

$$U_{g,t} = \alpha_g + \beta_g \times (U_t - U_t^*),$$

where the dependent variable is the unemployment rate of demographic group g (such as 25- to 34-year-old Hispanic women with some college), and the independent variable is the same economy-wide unemployment gap used above to measure business cycle conditions. The interpretation of this regression is similar: The variable α_g describes the average unemployment rate for a demographic group when the economy-wide unemployment rate is equal to the equilibrium rate, and β_g measures the cyclicality of the unemployment rate for that group.

Figure 4. Groups with Higher Unemployment Have More Cyclical Unemployment[a]

Source: Author's calculations, using the Current Population Survey.
a. Each point describes the results for a sex × age × education × race group, g; $U_{g,t} = \alpha_g + \beta_g \times (U_t - U_t^*)$.

I present the estimates from all 160 regressions in my figure 4, where each point shows the estimate of α_g and β_g for a specific demographic group. Because some of the cell sizes are quite small, the size of each point is proportional to the number of observations for that group in the underlying micro data.

Again, the finding is quite clear: Those narrowly defined demographic groups that tend to have higher unemployment rates (that is, higher αs) also tend to have more cyclical unemployment (higher βs). The data are clustered along a line of best fit that appears roughly linear, or perhaps slightly concave.

To my knowledge, this is a new finding in the literature, and it presents a stylized fact that I hope will be useful in guiding theoretical models of why different groups fare differently over the business cycle. It is consistent with the notion that a single index determines both an individual's average unemployment risk and the cyclicality of his or her unemployment rates. This could arise if demographic characteristics directly lead to both higher and more cyclical unemployment (an idea that Okun implicitly endorsed when he wrote that a high-pressure labor market might especially help particular groups, because that is "when market forces are making racial and sexual discrimination costly to employers"). It could also arise if these demographic characteristics are statistical proxies for some deeper notion of disadvantage or some other index relevant to labor market outcomes.

INTERNATIONAL EVIDENCE Aaronson and her colleagues are not just interested in whether disadvantaged groups have more cyclical unemployment. They see their unique contribution as testing the hypothesis that a high-pressure labor market may be different—perhaps even more effective—at improving the labor market outcomes of disadvantaged relative to advantaged groups. Effectively, they want to know whether the relative gains to disadvantaged groups from a "hot" labor market are even larger than might be expected from a linear relationship with the state of the aggregate economy. To study this, they examine how unemployment differentials—such as the difference between the unemployment rates of blacks and whites, or the difference between the unemployment rate of less educated workers and that of more educated workers—vary with the state of the business cycle. Previous research had shown that these differentials narrow when the national unemployment rate declines. They hypothesize that there might be an asymmetry to this cyclicality, so that the relationship between unemployment differentials and the state of the economy becomes stronger in hot economies.

The challenge is that by looking only at the United States, their sample includes just five episodes of "hot" labor markets (the late 1970s just before the Volcker disinflation; the late 1980s just before the 1990 recession; the middle to late 1990s, in the late stages of the Clinton-era boom; the middle to late 2000s, when unemployment barely dipped below the natural rate; and, finally, after 10 years of the recovery, following the Great Recession). Perhaps, then, it is unsurprising that their estimates of the extent (if any) of this asymmetry—shown in their tables 3 and 4—are extremely imprecisely estimated. Their estimates are sufficiently imprecise that they are left unable to reject the null hypothesis that this relationship is the same in both hot and cold labor markets, just as they are typically unable to reject the null that hot labor markets make unemployment differentials either 50 percent more or less sensitive to the state of the business cycle. My own conclusion is that their sample contains too few "experiments" of hot labor markets to be very informative about this issue. (The authors' subsequent attempt to parse the results by individual business cycle yields even lower power.)

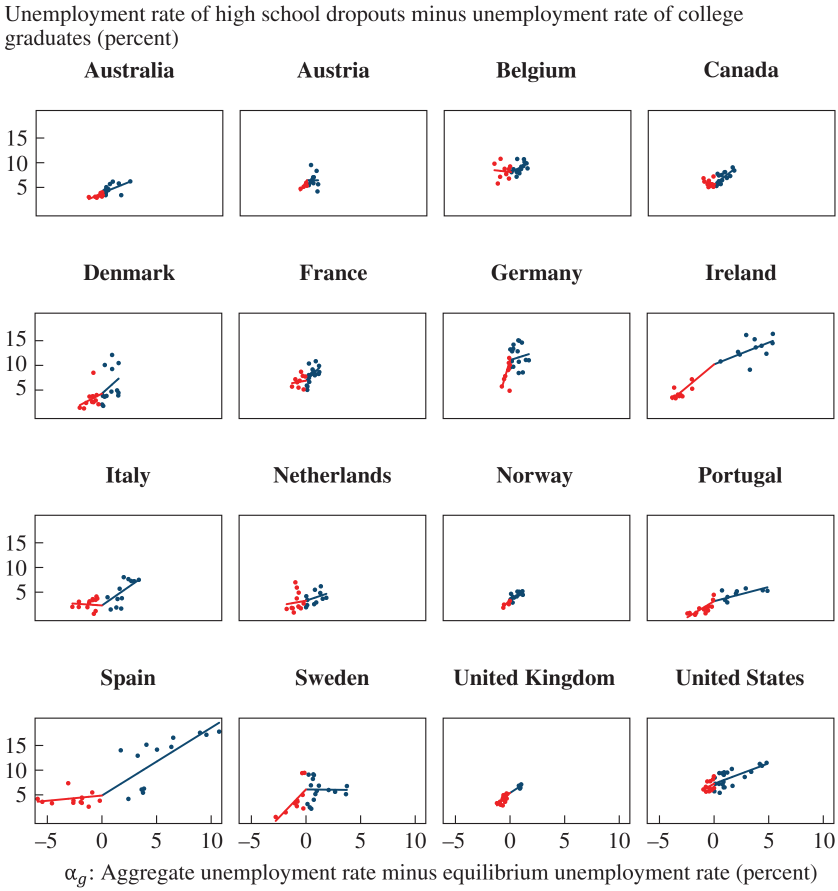
One natural response to underpowered results coming from a single-country study is to expand the sample to include the experiences of other countries. To this end, I collected unemployment data on those age 25–64 by educational attainment for nations that belong to the Organization for Economic Cooperation and Development (OECD). I drew these data from the *OECD Education Statistics* database and use them to construct unemployment differentials by education, focusing on the difference between the unemployment rate of those with less than an upper-secondary education ("high school dropouts") and those with a tertiary education ("college graduates"). I compare the evolution of each of this differential with the state of the national business cycle, as measured by the gap between a country's aggregate unemployment rate and the OECD's measure of that country's equilibrium unemployment rate, drawn from the May 2019 edition of its *Economic Outlook* database. This yields annual data covering up to 32 countries over the sample period 1981–2017 (albeit with some missing data).

My figure 5 illustrates how the educational unemployment differential—the unemployment rate of high school dropouts minus the (lower) unemployment rate of college graduates—varies with aggregate business cycle conditions. Though my full sample includes up to 32 countries, in order to keep the plots manageable, I show results only for 16 of the larger and more interesting OECD economies. In order to draw attention to hot labor markets, outcomes where the national unemployment rate is below that country's equilibrium unemployment rate are shown on the left of each panel in a lighter shade.

Two key facts are evident. First, there is a positive correlation between this unemployment differential and the state of the labor market. That is, the lower the national unemployment rate, the smaller are these education unemployment differentials. This finding is consistent with the earlier United States–centric literature that had found that the unemployment rate of disadvantaged groups is more sensitive to cyclical fluctuations. Indeed, across the 32 countries for which I have usable data, the "unemployment gap" is a statistically significant explanator of this unemployment differential at the 1 percent level for 20 countries, and at the 10 percent level for 28 countries. (The 4 countries where it is not significant all had fairly short samples.)

Second, there is not much evidence to support the hypothesis of Aaronson and her colleagues that this relationship steepens in a hot labor market. My figure 5 shows a regression line where I allow this relationship to change in a hot labor market (defined as one where the national unemployment rate is below the equilibrium unemployment rate). Across these

Figure 5. Unemployment Differentials by Education Level, Selected Countries



Source: Author's calculations, using *OECD Education Statistics and OECD Economic Outlook.*

countries, this relationship does not appear to systematically steepen (or flatten) in hot labor markets. Following their approach, I estimated regressions of the following form for each country:

$$U_{c,t}^{HS\ dropout} - U_{c,t}^{College} = \alpha_c + \beta_c\left(U_{c,t} - U_{c,t}^*\right) + \gamma_c\left(U_{c,t} - U_{c,t}^*\right) \times I\left(U_{c,t} < U_{c,t}^*\right).$$

In this case, γ_c measures how much this relationship steepens (or flattens) when the national unemployment rate falls below the equilibrium unemployment rate. Aaronson and her colleagues had hypothesized that γ_c would be positive. My point estimate was positive for 16 countries, and negative for 16 countries. Judged against a 10 percent significance level, it was significantly positive in 5 countries, significantly negative in 6, and insignificant in the remaining 21 countries. (And, in the absence of corrections for autocorrelation, this probably overstates significance.)

Finally, in order to allow the data to speak as clearly as possible, I pooled all the data into a country-year panel, to estimate one $\bar{\gamma}$ (rather than allowing it to vary by country)—while controlling for country fixed effects and allowing β_c to vary by country. This yielded a statistically significant negative coefficient, which is precise enough to reject the authors' hypothesis that it would be positive.

CONCLUSION My conclusion is that there is robust evidence across countries that a strong national labor market narrows unemployment differentials, but there is no support for the hypothesis of Aaronson and her colleagues that this relationship intensifies in a "hot" labor market.

From a policy perspective, I am not convinced that the authors' hypothesis is central to Okun's argument about the social benefit of pursuing full employment. Previous research has found—and this paper and my comment have confirmed—that hot labor markets help disadvantaged groups more than advantaged groups. As such, full employment reduces unemployment differentials between groups. The failure to identify an asymmetry does not undermine the broader point that full employment is valuable not only from an efficiency perspective but also because it yields an "equality dividend," in which labor market opportunity becomes more equally shared across demographic groups.

Finally, a personal note. The exploration of this question is one to which the late Alan Krueger contributed (see, for instance, the 1999 paper by Lawrence Katz and Krueger), and I know that he would have found this paper interesting. Alan was always deeply engaged by questions at the intersection of labor and macroeconomics, and he was a frequent and vigorous contributor to the Brookings Panel (and, indeed, in any setting where the issues of the day were debated). Our discussion of this work—as with our discussion of so many policy-relevant topics—was impoverished by his recent death. Alan, you are missed, not just by your colleagues who valued your insights but also by the less fortunate whose lives your work illuminated, and improved.

REFERENCES FOR THE WOLFERS COMMENT

Katz, Lawrence F., and Alan B. Krueger. 1999. "The High-Pressure U.S. Labor Market of the 1990s." *Brookings Papers on Economic Activity*, no. 1: 1–81. https://www.brookings.edu/wp-content/uploads/1999/01/1999a_bpea_katz.pdf.

Okun, Arthur M. 1973. "Upward Mobility in a High-Pressure Economy." *Brookings Papers on Economic Activity*, no. 1: 207–52. https://www.brookings.edu/wp-content/uploads/1973/01/1973a_bpea_okun_fellner_greenspan.pdf.

GENERAL DISCUSSION Katharine Abraham noted the importance of this paper by Stephanie Aaronson and her colleagues, especially the section examining whether there is evidence of positive hysteresis from running a tight labor market. Abraham observed that evidence of positive hysteresis would affect the trade-offs that policymakers face. On this note, she questioned the authors' decision to limit their analysis to people age 25–64 years, given that indications of negative hysteresis have been particularly apparent among those entering the labor market during a bad economic period. As such, investigating the subsequent effects of entering the labor market during a hot period might be especially useful.

Abraham commented that analyzing the effects of a hot labor market on the flows across labor market states, in contrast to the effects on stocks such as the unemployment rate and employment-to-population-ratio, would be especially interesting. She wondered whether running a hot labor market increases the likelihood of a person either finding a job or changing jobs. She imagined that labor market flow patterns are likely to differ by age.

Steven Davis reflected on a paper by Sherwin Rosen that highlights the dynamic complementarity between specific human capital investment and future utilization rates of that human capital.[1] Davis remarked that a person can invest in market-relevant and/or non-market-relevant skills As a person acquires market-relevant skills, for example, the reward to working in the market rises. There is also an effect in the other direction, whereby someone who anticipates future market work activity perceives a high return to acquiring market-relevant skills. In this sense, anticipated market work and the acquisition of market-relevant skills are mutually reinforcing. The same logic applies to nonmarket activities and the acquisition of skills that pay off in nonmarket activities. Davis commented that this phenomenon is highly relevant for thinking about potential mechanisms of a hot labor market because the causality is two-way and intertemporal. This two-way causality implies that isolating the effects of a hot labor market is empirically challenging.

1. Sherwin Rosen, "Specialization and Human Capital," *Journal of Labor Economics* 1, no. 1 (1983): 43–49.

Davis referred to a paper by David Neumark and Olena Nizalova that examines the connection between the employment rate of younger, less-educated people and whether the minimum wage in their local labor market was binding 10 years earlier.[2] Neumar and Nizalova find that younger, less-educated people who were affected by a binding minimum wage 10 years earlier will currently have lower employment rates. Davis remarked that this finding suggests that there are persistent effects of a person's ability to find work earlier in life on his or her likelihood of employment later in life. Thus, in considering the benefits of a tight labor market, the best place to look may not be the unemployment rate. Indeed, looking at other outcomes, such as employment and wages down the road, is likely to be especially enlightening. Davis noted that Rosen's logic also suggests important omitted variables—for example, the expected persistence of labor market tightness.

Davis noted that labor market tightness differs greatly across localities at a point in time. Spatial differences in tightness are potentially quite useful for estimating their effects on future labor market outcomes, despite the endogeneity of worker mobility across spatial labor markets.

Davis also argued that there is abundant evidence pointing to persistent positive effects of drawing less skilled people into the labor market, especially a hot one. For example, he described the research finding that tight labor markets facilitate job-to-job mobility. In addition, he noted research that links job-to-job mobility for younger workers with the ability to find a good-quality job match and get a larger share of the rents in the match. All in all, he asserted that economists have good reasons to believe that match quality improves the likelihood of higher wages, especially for less educated workers operating in a high-pressure environment. In this regard, he mentioned an empirical paper by Robert Topel and Michael Ward and a forthcoming theoretical paper by Gregor Jarosch that both study match quality effects and their connection to wages.[3]

Adele Morris wondered whether it would be possible to exploit variation across different sectors of the economy, given that some job sectors are more cyclical than others. This tendency for some sectors to be more

2. David Neumark and Olena Nizalova, "Minimum Wage Effects in the Longer Run," *Journal of Human Resources* 42, no. 2 (2007): 435–52.

3. Robert H. Topel and Michael P. Ward, "Job Mobility and the Careers of Young Men," *Quarterly Journal of Economics* 107, no. 2 (1992): 439–79; Gregor Jarosch, "Searching for Job Security and the Consequences of Job Loss," forthcoming, https://www.dropbox.com/s/whwpxtwskjzfq22/JobSecurity_121115_Full.pdf?dl=0.

cyclical than others implies that when the economy is hot, workers are disproportionately brought into the most cyclical sectors. As such, when the economy cools off, these same people are the most likely to experience unemployment. Thus, differentiating between workers—such as construction workers or schoolteachers—might give the authors additional insights.

Harry Holzer noted that the United States has experienced a large secular decline in the labor force participation rate among less educated males, which accelerated during the most recent economic downturn. He observed that examining the extent to which these males are returning to work and disaggregating some of the reasons they left work would be interesting. Moreover, he proposed that using a different data set, which includes people with criminal records or disability status, would likely be informative.

Holzer commented that research shows that supply-oriented interventions—like job training—are more effective when the economy is hot, especially job training centering on the tighter sectors. Importantly, this research also shows that sector-based training is more effective than other kinds of job training. Given this, Holzer pronounced the present moment to be a logical time to ramp up these investments and specifically target them toward workers who have permanently left the job market.

Jonathan Pingle emphasized Adele Morris's point about using industry data, noting that the excess cyclicality in the male unemployment rate in the last business cycle was likely due, in part, to the acute downturn in construction. Pingle also observed that structural sectoral shifts—such as the long-term decline in manufacturing jobs—could dampen or mask some of the cyclicality across some disadvantaged groups.

Susanto Basu linked the paper's discussion of employment and earning outcomes to the theme of persistence. Basu noted the concept of the "user cost of labor," introduced by Marianna Kudlyak, which is the difference between the present discounted wage and a point in time wage.[4] Cynthia Doniger disaggregated this concept of the user cost by education group, and finds that user costs are very procyclical.[5] Specifically, user costs are procyclical for the college educated, and are fairly procyclical for people

4. Marianna Kudlyak, "The Cyclicality of the User Cost of Labor," *Journal of Monetary Economics* 68 (2014): 53–67.

5. Cynthia L. Doniger, "Do Greasy Wheels Curb Inequality?" Finance and Economics Discussion Series 2019-021 (Washington: Federal Reserve Board of Governors, 2019), https://www.federalreserve.gov/econres/feds/files/2019021pap.pdf.

with some college education, but are not at all cyclical for those without a high school degree. Basu observed that the fact that the paper by Aaronson and her colleagues shows that less educated groups are the most likely ones to have a positive beta over the cycle for employment exactly because these same groups do not experience a positive beta for their wages, at least in the present value sense. Basu closed by noting that this phenomenon of positive gains in employment but not wages among the less educated also suggests that what is gained on the swings is lost on the roundabouts.

Ayşegül Şahin commented that it seems natural that employment gaps decrease when the economy is doing well. Agreeing with Abraham, Şahin observed that labor force attachment is important, and notes that she believes it is a positive development that attachment is rising faster for black males than white males. However, she pondered whether this progress was related to a hot economy or something else. For example, she noted that decreasing incarceration rates are also likely to be reducing labor market exits. Similar to the trend in the female labor force participation rate and the disappearance of the gender gap, the decreasing incarceration rate is increasing labor force attachment and reducing the unemployment rate. She concluded that examining the variation in socioeconomic factors could also shed light on whether these observed positive developments are actually related to the hot economy.

John Haltiwanger discussed the job-to-job flows (known as J2J) database, a new Census Bureau data product constructed using administrative data. J2J is an employer-employee matched database that tracks all businesses and all workers in the United States. In addition, J2J tracks characteristics about businesses and workers—such as age, race, ethnicity, and education—and is available at the level of metropolitan statistical areas (MSAs). These J2J data show—consistent with earlier evidence from the Current Population Survey—that the job ladder collapsed during the Great Recession. Furthermore, this collapse disproportionately affected the less educated, the young, and the disadvantaged. Haltiwanger noted that the slow recovery includes a slower recovery on the job ladder, especially among these groups. He concluded that he does not believe that $U - U^*$ (the unemployment rate minus the natural rate of unemployment) is the correct way to measure a hot labor market. Instead, he believes that a more general measure of labor market tightness would be useful, such as V (vacancies) over U.

David Romer cautioned about drawing policy implications about these issues without a macroeconomic model. In the extreme, effects at the macro level may undo or reverse the conclusions one might be tempted to draw based on intuition. If aggregate demand policy cannot affect the average unemployment rate and if aggregate welfare is linear in the unemployment rate, then the discovery that recessions are worse than perceived and booms are better than perceived has no implications for the welfare effects of stabilization policy. The reason is simply that under these assumptions, it is impossible to affect average welfare through stabilization policy. And if there are nonlinearities in the Phillips curve in the most plausible direction, with below-normal unemployment raising inflation more than above-normal inflation lowers it, then introducing volatility in the economy—by pushing unemployment below the natural rate, with a later period of unemployment above the natural rate to avoid a permanent increase in inflation—raises average unemployment over the cycle. In that case, a finding that unemployment is costlier than previously thought implies that such a policy reduces average welfare over the cycle, and that the welfare cost is higher than previously believed. In contrast, the opposite holds true if there are nonlinearities in the Phillips curve in the other direction. Similarly, nonlinearities in aggregate welfare as a function of the unemployment rate would also affect these calculations, and we have little evidence about such nonlinearities. In sum, Romer emphasized that it is the nonlinearities that drive any policy implications, rather than the first-order terms.

Robert Hall stated that his take on Kudlyak's work was different than Basu's. Kudlyak finds that there is an advantage to taking a job in a hot market relative to either what existing workers have or to the starting wage in a normal market. Moreover, this advantage is persistent over about six years. Hall noted that this finding does not directly relate to the question of a differential effect for disadvantaged people, but rather suggests that the effects are persistent. Hall affirmed that using a panel data set is important. In terms of wage rates, Hall referred to a large body of research that shows that though employment effects are not very persistent, wage effects are. To Hall, these findings indicate that looking at wage rates is the best way to explore the idea of persistence.

Robert Gordon reflected on the Okun coefficient, which up through the mid-1980s shows that unemployment moves about half as much as the gap between actual and potential output. This coefficient suggests that when unemployment goes down, output increases by 1 percent. This 1 percent bonus in a high-pressure economy comes from a combination of higher labor force participation, higher hours of work per employee, and higher productivity. Recent data, however, suggest a shift in Okun's law in the direction of a larger labor force response and a smaller productivity response. Indeed, Gordon indicated that recent data show very little procyclicality in productivity. Currently, the response of unemployment to a change in the output gap is more like 0.7 or 0.5, suggesting that the additional bonus of a hot labor market is not as large as it used to be. However, Gordon cautioned that, when thinking about the issue of the procyclicality of productivity, much of the change in the Okun coefficient is conditional on the jump of about 2.2 percent in productivity during the worst period of the Great Recession. To Gordon, this suggests that the increase in the sensitivity of unemployment to output change is partly an artifact of unusual behavior during the Great Recession, and that the issue of whether the economy got an additional bonus from higher productivity is still on the table.

Gilbert Metcalf considered a paper by Gordon Hanson, Chen Liu, and Craig McIntosh on changes in low-skilled immigration.[6] Metcalf wondered what the changes in low-skilled immigration mean for the findings of the paper by Aaronson and her colleagues. Metcalf also contemplated whether examining underemployment and its effects on income serves as a better measure for analyzing a hot labor market.

Stephanie Aaronson thanked both commenters, Julie Hotchkiss and Justin Wolfers, for bringing additional perspective to their work, and she also thanked all who participated in the general discussion. In response to the comments on exploiting regional variation, Aaronson noted that their paper does include an MSA analysis. This MSA analysis largely confirms the results in the rest of the paper: that there is weak evidence of a kink when the labor market gets hot, especially for African Americans. Similarly, Aaronson noted that the paper also includes an analytical comparison of metropolitan with nonmetropolitan areas. This analysis shows that the gap in labor market outcomes between metropolitan and nonmetropolitan areas is not sensitive to the aggregate unemployment rate, implying that what is going on in nonmetropolitan areas is recent, is structural, and does not seem to be affected by the business cycle.

Aaronson appreciated Abraham's question regarding younger workers, especially because this age group was an original focus of Okun. Aaronson remarked that although their paper does not show results for younger workers, earlier versions of the paper did include those age 16–65 and

6. Gordon H. Hanson, Chen Liu, and Craig McIntosh, "The Rise and Fall of U.S. Low-Skilled Immigration," *Brookings Papers on Economic Activity*, Spring 2017, 83–151, https://www.brookings.edu/bpea-articles/along-the-watchtower-the-rise-and-fall-of-u-s-low-skilled-immigration/.

that these results showed more evidence of hysteresis. She mentioned that exploring whether there is evidence of positive hysteresis among younger workers would be easy to do, and the authors hope to present these new results in the final version of the paper.

In response to Davis, Aaronson agreed that using panel data is valuable. Aaronson noted a study by Julie Hotchkiss and Robert Moore that explores the concept of positive hysteresis using panel data from the National Longitudinal Survey of Youth.[7] Aaronson agreed that employment effects are likely to be short-lived, whereas wage effects may last longer for some groups of workers.

Aaronson expressed concern about using industry-level data, noting the difficulty of attaching people to an industry over time, but she agreed that some workers are brought into more cyclical industries. On this note, she mentioned that there is also evidence that finds no indication of hysteresis for precisely this reason: workers who are brought into highly cyclical industries are the same workers who become unemployed again when the economy turns cold.

Aaronson thanked David Romer for his comments, and she reflected on a paper that he and Christina Romer wrote for the 1998 Economic Policy Symposium at Jackson Hole that discussed the trade-offs of running a hot economy.[8] Aaronson remarked that running a hot economy can be a powerful tool to help disadvantaged workers; however, it remains unclear whether a hot economy can be run for long enough to provide a substantial benefit to these same workers. Aaronson explained that neither her and her colleagues' paper nor historical experience provides an indication of a long-term benefit. She cautioned that policymakers should be careful about using policy levers—such as running a hot economy—to solve structural problems. Such problems, she advised, are likely to require more than monetary policy to solve.

William Wascher commented on Haltiwanger's warning that using $U - U^*$ might not be the best measure of the tightness of the labor market.

7. Julie L. Hotchkiss and Robert E. Moore, "Some Like It Hot: Assessing Longer-Term Labor Market Benefits from a High-Pressure Economy," Working Paper 2018-1 (Atlanta: Federal Reserve Bank of Atlanta, 2018), https://www.frbatlanta.org/-/media/documents/research/publications/wp/2018/01-assessing-longer-term-labor-market-benefits-from-a-high-pressure-economy-2018-01-30.pdf.

8. Christina D. Romer and David H. Romer, "Monetary Policy and the Well-Being of the Poor," paper presented at Economic Policy Symposium on Income Inequality Issues and Policy Options, sponsored by Federal Reserve Bank of Kansas City, Jackson Hole, Wyo., August 27–29, 1998, https://www.kansascityfed.org/publicat/sympos/1998/S98romer.pdf.

Wascher noted that previous iterations of his and his colleagues' paper had included a model that picked the unemployment rate where the kink might occur. This model chose an unemployment rate that was often higher than U^*. As such, thinking more broadly about how to carefully measure labor market tightness would be useful. Wascher agreed with Metcalf's comment that underemployment is likely to be an issue. Wascher noted that the paper attempts to get at this underemployment question through their analysis of wages, but he agreed that this problem requires more attention.

GILBERT E. METCALF
Tufts University

On the Economics of a Carbon Tax for the United States

ABSTRACT Climate change is driven by the buildup of greenhouse gases (GHGs) in the atmosphere, which is predominantly the result of the world's consumption of fossil fuels. GHGs are a global pollution externality for which a global solution is required. I describe the role a domestic carbon tax could play in reducing U.S. emissions and compare and contrast alternative approaches to reducing our GHG pollution. Carbon taxes have been implemented in 23 jurisdictions around the world. I provide evidence on emission reductions and the economic impact of British Columbia's carbon tax, a broad-based carbon assessment that has been in effect for over a decade. I also provide an analysis of carbon taxes used in the countries that belong to the European Union.

Climate change is a classic global pollution externality, with billions of polluters creating damage for billions of people. Moreover, the world's continued use of fossil fuels and other GHG-emitting activities creates damage that will affect future generations. This paper considers the role that a carbon tax could play in the United States as its contribution to reducing emissions. Although climate change is a global problem and the United States has been surpassed by China as the world's largest emitter, I focus on domestic policy. A domestic carbon tax alone will not make a major dent in global emissions. But it is difficult to imagine other

Conflict of Interest Disclosure: The author is a professor of economics at Tufts University and a research associate at the National Bureau of Economic Research. Beyond these affiliations, the author did not receive financial support from any firm or person for this paper or from any firm or person with a financial or political interest in this paper. He is currently not an officer, director, or board member of any organization with an interest in this paper. No outside party had the right to review this paper before circulation. The views expressed in this paper are those of the author, and do not necessarily reflect those of Tufts University or the National Bureau of Economic Research.

countries taking aggressive action to curb GHGs if the United States does not enact strong policy measures to reduce its emissions.

This paper argues that a carbon tax should be the central element of U.S. policies to reduce emissions. Putting a price on carbon pollution is a straightforward application of Pigouvian pollution pricing and a textbook response to the market failure arising from pollution. Although a carbon tax is a necessary element in a cost-effective policy approach to pollution, it is not sufficient. Moving to a zero-carbon economy will require new inventions and production processes. And research and development (R&D) will be key to their successful diffusion—whether it is advanced battery storage, carbon capture and storage, or inexpensive, safe, and modular nuclear power. Information and new knowledge are pure public goods that are underprovided in a market economy.[1] The information market failure is a general market failure and not one specific to GHGs. But R&D is central to any solution to the GHG problem, and directed R&D support can ensure that emission reduction targets are met with lower carbon tax rates and the consequent economic costs of the tax, a point made by Daron Acemoglu and others (2012) and by Acemoglu and others (2016). These two market failures—pollution and the pure public goods nature of R&D—should drive our choice of policy. In section V, I discuss other policy needs to complement the carbon tax and energy-related R&D.

Section I of the paper briefly describes climate change and the damage from failing to act to reduce U.S. carbon pollution. Section II compares and contrasts a carbon tax with alternative policy approaches. In section III, I survey the use of carbon taxes around the world. In section IV, I present some evidence on the economic impact of carbon taxes, with a particular focus on the emissions and GDP effects of British Columbia's carbon tax. Section V presents thoughts on policy design, and section VI concludes.

I. Climate Change

"Climate change" is a catchall term for the climate effects arising from accumulations of GHGs in the Earth's atmosphere. The most prominent GHG is carbon dioxide (CO_2), which accounts for over three-quarters of

1. There are two issues here. First is the ability of private inventors to appropriate the benefits of their inventions. Patent protection is an imperfect policy tool for this, thereby deterring R&D. Second is the fact that even with the ability to fully appropriate the gains, the pure public goods nature of new ideas means that the social gains likely exceed the private gains.

global emissions. Methane is the second most prominent GHG, accounting for a further almost 16 percent of global emissions. Nitrous oxides (N_2O) and other gases account for the remaining close to 8 percent of GHG emissions. CO_2 is a higher share of U.S. GHG emissions, accounting for about 82 percent, with methane accounting for about 10 percent and N_2O and other gases accounting for the remaining close to 8 percent.[2]

Focusing on sectors, about 84 percent of U.S. GHG emissions are in the energy sector. Agriculture accounts for about 9 percent, industrial processes and product use for about 6 percent, and waste for about 2 percent. Within energy, about 94 percent of emissions are from CO_2, of which about 97 percent is associated with fossil fuel combustion. Breaking down energy-related fossil fuel combustion CO_2 emissions, about 36 percent are from transportation, about 16 percent industrial, about 11 percent residential and commercial, and 36 percent from electricity.[3]

The damage from GHG emissions stem from the stock of these gases in the atmosphere. Central to understanding the effect of accumulating stocks of CO_2 in the atmosphere on climate change is a scientific parameter known as equilibrium climate sensitivity. Equilibrium climate sensitivity measures the long-run equilibrium increase in temperature arising from a change in the stock of GHGs in the atmosphere. Just as the glass roof of a greenhouse traps solar radiation and raises the temperature inside the greenhouse, CO_2 and other GHGs trap solar radiation in our atmosphere and raise the planet's temperature. Hence the reference to "greenhouse gases" and the greenhouse effect of climate change. How fast the temperature rises in response to an increase in the stock of GHGs in the long run depends on the climate sensitivity parameter.[4]

Over one hundred years ago, Sweden's Svante Arrhenius, a childhood mathematics prodigy and Nobel Prize–winning chemist, made the first estimates of climate sensitivity in his 1906 book *Worlds in the Making*.

2. These data are for 2014 and are taken from the World Resources Institute's CAIT Climate Data Explorer (cait.wri.org). Emissions of non-CO_2 gases are converted to a CO_2 equivalent using a 100-year global warming potential taken from the 1996 *Second Assessment Report* of the Intergovernmental Panel on Climate Change.

3. These are shares of total GHG emissions as reported in U.S. Environmental Protection Agency (EPA 2018, tables 2-3, 2-4, and 2-5). Shares do not account for any forest or land use sinks. Electricity is used by the other sectors. If attributed to those sectors, the residential and commercial sectors would tie with transportation as the most carbon-intensive sectors (about 36 percent each).

4. Equilibrium climate sensitivity measures the long-run equilibrium response. Transient climate response measures the temperature response over a shorter period. Figure 1 shows the relationship between carbon concentrations and temperature increase that reflects the transient climate response relationship.

He estimated the value of the climate sensitivity parameter to be 4 degrees Celsius—that is, a doubling of GHGs leads to an increase in temperature by 4 degrees Celsius (just over 7 degrees Fahrenheit). He made this calculation notwithstanding the very early state of climate science and the lack of current, let alone historical, data on temperature and GHG concentrations. His estimate of climate sensitivity is remarkably durable. Despite the complexity of modeling climate sensitivity, modern estimates are in the ballpark of Arrhenius's hundred-year-old estimate.

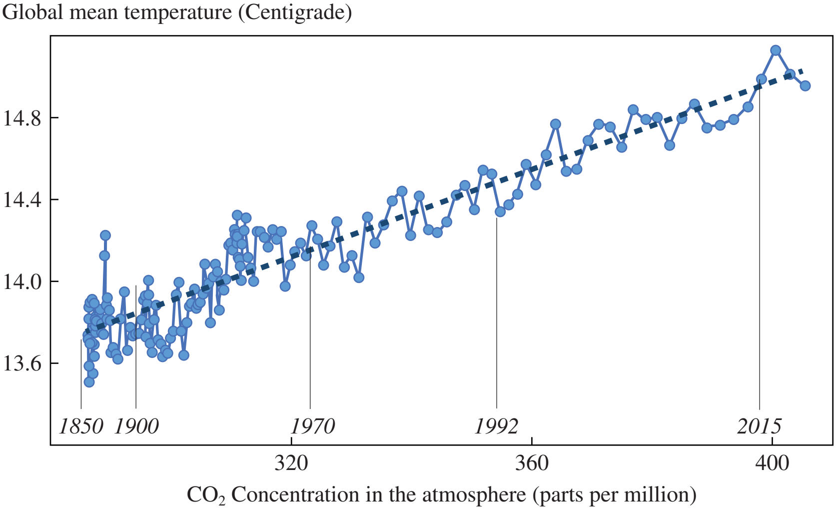
Pre–industrial era concentrations of CO_2 in the atmosphere are typically pegged at 280 parts per million, though air samples taken from Antarctic ice cores make clear that concentrations have ranged between 180 and 290 parts per million over the past 400,000 years (Petit and others 1999). Current measurements of CO_2 have been taken on a continuous basis in Hawaii starting in 1958, when Charles Keeling installed monitoring equipment on the upper slopes of the Mauna Loa volcano, which are just over 11,000 feet above sea level. The Keeling Curve shows a dramatically rising concentration of CO_2 in the atmosphere, with current monthly average concentrations topping 405 parts per million. Figure 1 shows the relationship between atmospheric CO_2 concentrations and global mean temperatures since 1850.

The U.S. National Oceanic and Atmospheric Administration publishes its Climate Extremes Index as a way to summarize extreme temperature (high and low), precipitation, droughts, and tropical storm intensity with data going back to 1910. Six of the top 10 extreme climate years have occurred since 2005, and each of the years since 2015 has been among the top 6 extreme years.[5] This index highlights the fact that climate change is as much (if not more) about climate variability than it is about warming.

Below, I discuss the economic costs of climate policy. Any discussion of policy costs should recognize that failing to act also has costs. Although a detailed analysis is beyond the scope of this paper, a few comments are in order. Until recently, most measures of the damage from GHG emissions were derived from reduced-form damage functions embedded in integrated assessment models, such as the Nordhaus Dynamic Integrated Climate-Economy Model. William Nordhaus (2013) describes the various cost factors and models damage (as a percentage of global output) as an (approximately) quadratic function of temperature increase. In a recent

5. The data for the National Oceanic and Atmospheric Administration's Climate Extremes Index for the contiguous United States are published at https://www.ncdc.noaa.gov/extremes/cei/graph/us/cei/01-12.

Figure 1. The Relationship between Atmospheric CO_2 Concentrations and Global Mean Temperatures since 1850



Source: CO_2 data are taken from Antarctic ice core samples (pre-1958) and the Keeling data, as reported at http://scrippsco2.ucsd.edu/data/atmospheric_co2/icecore_merged_products. Global mean temperatures are from Berkeley Earth, at http://berkeleyearth.lbl.gov/auto/Global/Land_and_Ocean_summary.txt. The format of this figure is due to Robert Rohde of Berkeley Earth. A linear regression of the change in temperature from 1850 on the log of the ratio of CO_2 concentrations since 1850 yields an estimated 2.5-degree Celsius increase in temperature from a doubling of CO_2 concentrations. This regression fit is more akin to the transient climate response than the equilibrium climate sensitivity.

meta-analysis, Nordhaus and Andrew Moffat (2017) find no evidence for sharp convexities or discontinuities in the damage function, and they find damage on the order of 2 percent of global income for a 3-degree C increase in temperature and 8 percent at 6 degrees C. They caution, however, that damage estimates are not comprehensive and, in some areas, are little more than guesswork. As a result, these damage estimates should be viewed as lower bounds.

Solomon Hsiang and others (2017) construct detailed estimates of the damage from climate change in the United States at the county level, and they find that the combined market and nonmarket damage for a 1-degree C increase in temperature is on the order of 1.2 percent of GDP. Damage is unequally distributed, with higher damage in southern areas. By the end of this century, they estimate that the poorest third of U.S. counties have a 90 percent chance of experiencing damage between 2 and 20 percent of county income in a business-as-usual scenario with no action to reduce emissions.

The cost of climate change includes both damage and the costs of adaptation. As temperatures increase, we can expect to see greater penetration and use of air conditioners—a form of adaptation. Infrastructure investments to cope with more frequent and severe storms of a Sandy type are also forms of adaptation. Adaptations, of course, come with their own costs. The International Energy Agency (2018b) estimates that household ownership of air conditioners will rise from 1.1 billion units in 2016 to over 4 billion units by 2050. The electricity needed to power those new air conditioners exceeds the current electricity consumption in Germany and the United States.

II. Theory

Policymakers have a variety of instruments at their disposal to bring about a reduction in GHG emissions. They can raise the cost of emissions, lower the cost of clean alternatives to fossil fuels and other GHG sources, and impose regulations mandating specific technologies or benchmarks for emission reductions, among other options. In this section, I compare and contrast the various alternatives and argue that a carbon tax is the most cost-effective way to achieve a given reduction in GHG emissions.

II.A. Putting a Price on Pollution

Arthur Pigou is credited with the idea of using taxes to correct the market failure arising from the presence of externalities, as explained in his 1920 book *The Economics of Welfare*. The problem with pollution is that there is a divergence between the private and social costs of a good due to pollution, with the divergence equal to the marginal damage from the pollution. If this is the problem, argued Pigou, then taxing the pollution at its social marginal damage would equate private and social marginal costs and ensure an efficient market outcome.

For many pollutants, taxing the pollution is difficult if not impossible, whereas taxing the good associated with the pollution is more practical. Such is not the case, however, for energy-related CO_2 emissions. The amount of CO_2 associated with burning a ton of coal, a gallon of gasoline, or a therm of natural gas is, for all intents and purposes, constant.[6] Changes

6. Different grades of coal release different amounts of CO_2 per ton burned. But the differences are well understood and limited in number, making it straightforward to apply a carbon tax to coal either at the mine mouth or at the site where burned—or anywhere in between.

in industrial processes may affect the amount of fossil fuel burned but not the emissions per unit of fuel input.[7]

A Pigouvian tax is especially attractive in a situation where it is (relatively) easy to measure the marginal damage from the pollutant but where it is difficult to identify the individuals suffering the damage from pollution. In such an instance, bargaining between the polluter and those affected by pollution, à la Ronald Coase, cannot substitute for government intervention. Coase (1960, 852) understood this: "In the standard case of a smoke nuisance, which may affect a vast number of people engaged in a wide variety of activities, the administrative costs might well be so high as to make any attempt to deal with the problem within the confines of a single firm impossible. An alternative solution is direct government regulation."

Put differently, Coasian bargaining requires reasonably low transaction costs (along with clear property rights) for private bargaining to substitute for government intervention. Climate change has especially high transaction costs given the number of people affected, both across countries and across time.

A Pigouvian tax is a market-based instrument to control pollution, in the sense that it allows the market to operate once prices have been adjusted through the use of a Pigouvian tax. A cap-and-trade system is an alternative way to set a price on pollution. Whereas a carbon tax puts a price on CO_2 pollution and lets the market determine the amount of pollution, a cap-and-trade system puts a cap on pollution and lets a market operate in the buying and selling of rights to pollute (subject to the cap) and so determine a market clearing price. The earliest significant cap-and-trade system was the Acid Rain Program, which was established as part of the Clean Air Act Amendments of 1990.[8] The European Union's Emission Trading System (ETS) is the largest GHG cap-and-trade system established to date (World Bank Group 2018). The cap-and-trade concept is credited to the Canadian economist John Dales (1968) and builds on Ronald Coase's conception of the pollution problem as one of incomplete property rights (Coase 1960). By establishing a cap on pollution and distributing rights to pollute, a cap-and-trade system establishes clear (albeit limited by the cap) property rights to pollute.

7. The one major exception is carbon capture and storage, where CO_2 is captured when the fuel is burned and permanently stored to prevent its release into the atmosphere. I discuss carbon capture and storage and its treatment under a carbon tax in section V.

8. Schmalensee and Stavins (2013) provide a history and assessment of the Acid Rain Program.

An extensive literature compares and contrasts a carbon tax and a cap-and-trade policy. Although the economic literature suggests that a carbon tax is more efficient ex ante than cap and trade in a world with uncertain marginal abatement costs, the relative efficiency of the two instruments depends on underlying modeling assumptions.[9] The efficiency differences between traditional regulation and a market-based instrument like a carbon tax or cap-and-trade system are likely to be much greater than the differences between the latter two policies.[10]

Setting aside economic efficiency, three factors favor carbon taxes over cap-and-trade systems.[11] First, a cap-and-trade system fixes emissions but allows prices to vary as market conditions change. This can lead to price volatility and uncertainty for firms planning long-lived, capital-intensive projects. The Acid Rain Program illustrates the potential for price volatility. Allowance prices fluctuated anywhere from zero to $1,200 in the five years between 2005 and 2010.[12] Price fluctuations are not limited to the Acid Rain Program of the U.S. Environmental Protection Agency (EPA). Allowance prices in the European Union's ETS fell by one-third in one week in April 2006 and by a further 20 percent over the next month upon release of information that initial allowance allocations had been too generous.[13]

The second difference between the two policy instruments is in administrative complexity. The United States has a well-developed tax collection system, including systems in place to collect taxes on most fossil fuels. A cap-and-trade system, in contrast, requires an entirely new administrative structure to create allowances, track them, hold auctions or otherwise distribute them, and develop rules to avoid fraud and abuse. Fraud is a particularly significant problem in a system that is creating brand-new assets (emission allowances) worth billions of dollars. This is not just a

9. The literature comparing efficiency of the two instruments draws heavily on the seminal paper of Weitzman (1974). Weitzman's paper considered a flow pollutant. Papers that extend the Weitzman framework to consider a stock pollutant like GHGs include Hoel and Karp (2002), Newell and Pizer (2003), Karp and Zhang (2005), and Karp and Traeger (2018), among others. Excepting the last paper, the papers tend to favor a price instrument (tax) in the presence of a stock pollutant. Note, too, that the Weitzman framework assumes a once-and-for-all decision on a cap or tax schedule. If updating is possible, the differences between the two instruments shrink, if not disappear.

10. Carlson and others (2000) suggest that the cost of regulating sulfur dioxide emissions with a cap and trade could be reduced as much as one-half compared with traditional command-and-control regulation. See also Ellerman and others (2000).

11. I elaborate on these issues in Metcalf (2019). Goulder and Schein (2013) have a similar list.

12. See Schmalensee and Stavins (2013, figure 2).

13. The price decline is discussed in Metcalf (2009).

theoretical concern. In January 2011, the EU had to suspend trading in allowances when $9 million of allowances were stolen from an account in the Czech Republic. The EU commissioners noted that hackers had also broken into accounts in Austria, Poland, Greece, and Estonia and that as much as $40 million in allowances was stolen.[14] Though tax evasion is certainly a potential problem, the United States has a strong culture of tax compliance. The risk of cybertheft from electronic registries in a cap-and-trade system is likely to present a greater problem than the risk of tax evasion in a carbon tax.

The final difference between a carbon tax and a cap-and-trade system is the potential for adverse policy interactions that can work against the goal of reducing emissions. This is a big problem for cap-and-trade systems. Consider a cap set with a goal of realizing allowance prices of $40 a ton. This price target would contribute to driving innovation and the development of new carbon-free technologies that we will need to get to a zero-carbon economy by the end of the century. Investors will not place risky bets on new energy technologies that reduce emissions unless they can be confident that there is a good chance of earning a high return on this investment. The higher the carbon price, the more confident they can be that their investment will earn a return that will pay for the risk they will be taking. This is because a high carbon price drives up the cost of natural gas, petroleum, and coal, and can make a new zero-carbon investment competitive in the market, even at a cost that is high enough to repay the investors for the risks they took in underwriting a new and unproven technology.

Any additional policies enacted to reduce emissions in sectors covered by the cap-and-trade program (for example, low carbon fuel standards or renewable portfolio standards) will do nothing to reduce emissions but can only undermine allowance prices in the program. Any emission reductions in these supplementary programs will simply be offset by increases in emissions elsewhere, assuming the cap is binding. All that can happen is that the allowance price falls as the cap is loosened.

This is precisely what has happened in the major cap-and-trade programs. They have all struggled to set a price at a level that drives significant reductions in carbon pollution. Since trading began in 2013 for the current phase of the European Union's ETS (2013–20), prices have generally ranged between $3 and $8 per ton and only broke through the

14. The cybertheft story is reported by Chaffin (2011) and Lehane (2011), among others.

$10 barrier in March 2018. Prices in the earlier trading period (2008–12) were not much higher. When allowances for this commitment period were first issued, prices rose to nearly $36 a ton but quickly fell by about half and subsequently drifted down.[15]

To address low prices in the ETS, the EU initiated a program to reduce a surplus of allowances in the system that stemmed, in part, from the 2008 recession. The EU will reduce the surplus by one-quarter each year between now and 2024 by adding the allowances to its Market Stability Reserve.[16] This has helped raise ETS allowance prices to their current level (as of July 2019) of about $30 a metric ton.[17]

The World Bank's 2018 annual review of carbon pricing tracks carbon pricing in roughly 40 countries and 20 cities, states, and regions. The highest carbon price among the cap-and-trade systems surveyed in the review is about $16 a ton. In contrast, 5 countries have carbon tax rates of at least $50 a ton, with Sweden leading the group at about $140.

The most powerful arguments in favor of cap-and-trade programs over carbon taxes are that (1) prices are not being set directly by politicians, and so political distance is created for risk-averse policymakers; and (2) allowances created in a cap-and-trade program are valuable assets that policymakers can distribute in ways to reduce political opposition. For example, the Acid Rain Program created roughly 10 million allowances in 2000. With an average spot price of just under $145 a ton, the allowances disbursed that year were worth $1.45 billion. The Acid Rain Program distributed allowances for free to owners of coal-fired power plants based on their historic coal use. This certainly eased opposition to the program. Using allowances to overcome opposition was behind the complex allocation process in the American Clean Energy and Security Bill (HR 2454),

15. Allowance prices for the 2013 period forward are taken from the European Energy Exchange website (https://www.eex.com/en/market-data/environmental-markets/spot-market/european-emission-allowances). Prices from the 2008–12 period are from Koch and others (2014). Euro prices are converted to dollars at the rate of $1.15 per €1, the exchange rate as of January 10, 2019.

16. The announcement of allowances in circulation was published at https://ec.europa.eu/clima/news/ets-market-stability-reserve-will-start-reducing-auction-volume-almost-265-million-allowances_en. Also see Lewis (2018). Rules for adding allowances to or withdrawing from the EU's Market Stability Reserve were established in 2015 to go into operation in 2019. As of May 2018, the EU estimated that over 1.6 billion allowances were in circulation. Allowances in excess of 833 million are deemed surplus and subject to being added to the Market Stability Reserve.

17. A similar problem bedevils the Regional Greenhouse Gas Initiative, a cap-and-trade system for electricity in the U.S. Northeast (Metcalf 2019).

the cap-and-trade law passed by the U.S. House of Representatives in 2009 that ultimately failed in the Senate. A free allowance allocation can help grease the political wheels and contribute to passage of cap-and-trade legislation. But this is very expensive grease! The Congressional Budget Office estimated that the value of the free allowances in that bill would be nearly $700 billion over a 10-year period.[18]

Giving allowances to polluting firms for free raises important distributional questions. Giving firms $700 billion in free allowances has the same effect on their bottom line as giving them cash. The result is a windfall for shareholders—profits and share prices go up. This is what happened in Europe when the European Union set up its CO_2 cap-and-trade program and gave allowances to the firms that were subject to the cap.[19] Whether this is fair is a matter of debate. But the very complexity of the cap-and-trade approach means that the public did not really understand the massive transfer taking place in the EU's ETS or that would have taken place if the U.S. cap-and-trade legislation had gone into effect.

II.B. Regulation

Although the focus above has been on market-based instruments, the reality is that most of the polices to address climate change rely on various forms of regulation, subsidies, and voluntary actions or information. The two most important regulations that have been put forward to address GHGs at the U.S. federal level are the corporate average fuel economy (CAFE) standards and the regulation of CO_2 emissions in the power sector under the Clean Air Act. Recall that transportation and electricity generation each accounted for about 36 percent of energy-related CO_2 emissions in 2016. These two regulatory targets thus account for nearly three-quarters of these emissions.

After the U.S. Supreme Court ruled in 2007 that GHGs were air pollutants that could be regulated under the Clean Air Act, the EPA in 2009 issued an endangerment finding determining that GHGs should be subject to regulation and began the process of promulgating regulations. Numerous papers have been written on the relative inefficiency of fuel economy regulation relative to a Pigouvian tax—see, for example, the recent review by

18. Congressional Budget Office Cost Estimate of HR 2454, June 5, 2009 (https://www.cbo.gov/publication/41189).

19. Smale and others (2006) examine five energy-intensive sectors in the United Kingdom and conclude that profits in most of the sectors rise following the imposition of a cap-and-trade system with free allowance allocation.

Soren Anderson and James Sallee (2016). Taxes on emissions—for transportation, this can be translated into a tax on gasoline use—create incentives for consumers to purchase more fuel-efficient vehicles, drive fewer miles in the aggregate, and scrap fuel-inefficient vehicles sooner. A fuel economy standard mandating that an automaker's vehicle fleet must meet minimum fuel economy standards in toto also incentivizes the purchase of more fuel-efficient vehicles. But the higher fuel economy drives down the cost of driving per mile and thus can lead to more driving—the rebound effect. Moreover, fuel economy standards only apply to new vehicles. This increases the value of fuel-inefficient vehicles already on the road and delays their eventual scrappage, an effect first pointed out by Howard Gruenspecht (1982). All in all, these factors lead to fuel economy standards being less cost-effective than an emissions tax for achieving given emission reductions. Valerie Karplus and others (2013), for example, find that fuel economy standards are 6 to 14 times more expensive than a fuel tax to achieve the same emission reductions.[20] Mark Jacobsen (2013) finds CAFE is a little over three times the cost of a gasoline tax per ton of CO_2 avoided in a model where technology can respond to the mandate or higher fuel costs.

The Obama administration imposed tighter fuel economy standards for cars and light trucks for model years 2022–25 that would have raised the fleetwide average to 54.5 miles per gallon for 2025. This essentially would double fuel economy from the model year 2011 fleet standards of 27.3 miles per gallon.[21]

In August 2015, the Obama administration released the Clean Power Plan, a set of EPA regulations to cut GHG emissions from existing electric power plants.[22] The plan used building blocks of potential emission reduction channels—including efficiency improvements in boilers, generation shifting (from emissions-intensive fuel sources to less intensive sources), and increased generation from new low- or zero-emitting sources. Based on the EPA's analysis of the potential for emission reductions in each state, targets were set that could be in the form of emission rate standards,

20. Federal policy also includes various tax provisions that create an explicit or implicit tax on fuel economy. Sallee (2011) reviews these and notes that the inefficiency is exacerbated by gaming that results from the way the taxes are designed.

21. *Federal Register* 74, no. 59: 14196–556. The model year 2022–25 standards are described by NHTSA (2011).

22. The final plan was published in *Federal Register* 80, no. 205 (October 23, 2015): 64661–65120.

mass-based standards, or a "state measures" standard. States could also join together to create a regional cap-and-trade program, which, in the limit, could mimic a national cap-and-trade program for the electricity-generating sector. All this is moot, however, because then–EPA administrator Scott Pruitt issued a proposed rule to repeal the Clean Power Plan in October 2017 (Eilperin 2017). Because the endangerment finding is still in place, the EPA is required to propose a new rule. We can expect litigation no matter what approach the Trump administration takes to water down if not eliminate GHG regulations for the power sector.

The CAFE regulations and the Clean Power Plan illustrate the political vulnerability that results from using regulation to advance mitigation goals. In August 2018, the Trump administration announced a reworking of the model year 2022–25 standards as the Safer Affordable Fuel-Efficient (SAFE) Vehicles Rule that would freeze fuel economy standards at model year 2020 levels through model year 2026 (NHTSA 2018). States are currently fighting this rule rollback in the courts. And, as noted above, the Trump administration is working to eliminate the Clean Power Plan. Executive action using regulatory authority is subject to the political risk of changes in administration that can lead to a new reading of laws and consequent changes in enforcement and stringency. Meanwhile, opponents of the rule changes (whether made by the Obama or Trump administration) have challenged the changes in the courts, thereby adding to the policy risk and uncertainty.

II.C. Subsidies

Subsidizing activities that compete with the polluting activity can reduce pollution and is particularly attractive to politicians. After all, subsidies generally lower costs for their constituents. The problem, however, is that someone has to pay for the subsidy. These costs, in general, are spread across many people; so though the aggregate cost of the subsidy might be large, the cost to any individual may be too small to notice.

Renewable Portfolio Standards (RPSs) are common policies at the state level. RPS programs are a blend of regulation and subsidy and are currently in place in 29 states (North Carolina Clean Energy Technology Center 2018). An RPS policy mandates that a certain fraction of the electricity sold in the state must come from a designated renewable source, such as wind or solar. Massachusetts, for example, has a requirement that every private company selling electricity in the state in 2020 must prove that it has satisfied its 15 percent RPS obligation. Companies demonstrate compliance by submitting renewable energy credits (RECs) to the state each year. RECs are like vouchers that the state gives to renewable electricity producers for every megawatt-hour (1,000 kilowatt-hours) of electricity the renewable facility generates. The owners can then sell those vouchers to electricity distribution companies that buy as many RECs as they need to comply with the state law. The payment from the company that sells electricity to retail customers is made over and above the payment for the electricity that the renewable generator sells into the system. An owner of a commercial solar farm selling electricity into the grid might get paid between 2 to 10 cents per kilowatt-hour, depending on the time of day the power is sold. The owner could also sell a REC to a utility that needs it to comply with the RPS rule. This might bring another 25 to 28 cents per kilowatt-hour (based on solar REC prices in 2014 in Massachusetts). The cost of the REC gets folded in to the cost of generation and passed on to ratepayers.

Although the REC costs get passed on to ratepayers, the cost increase is blunted to some extent by the fact that wind and solar power have very low (essentially zero) operating costs. As a result, electricity prices do not go up as much as when a tax is imposed. Keeping prices down discourages firms and individuals from investing in energy efficiency to reduce consumption. And though a tax increase may be unpopular, it does raise revenue that could be returned to taxpayers in a way that preserves the energy-saving price signal while also offsetting the income loss from higher electricity rates. Blunting the price signal raises the cost of RPS emission reductions relative to a carbon tax. A recent study found that the cost of cutting carbon emissions in the electricity sector by 10 percent was over six times higher with an RPS program than with a carbon tax applied to fuels used to generate electricity.[23]

Rather than have the ratepayer pay for the subsidy, as in RPS programs, taxpayers could finance it. Since the first energy crisis back in the 1970s, Congress has provided tax breaks to encourage various energy technologies, including breaks for developing and using renewable technologies.[24] Historically, the biggest tax breaks have been tax credits for projects that generate electricity from solar, wind, geothermal, or other

23. Reguant (2018) carries out the study comparing RPS and carbon taxes in the electricity sector. Fischer (2010) has shown that RPS programs can actually reduce electricity prices because the price of wind or solar at the margin is zero in contrast to natural gas, which, while cleaner than coal, still has a cost at the margin.

24. Since the inception of the tax code, there have been large tax breaks for domestic oil and gas drilling. Metcalf (2018) shows that these incentives have had modest effects on domestic oil and gas production but are costly to the U.S. Treasury.

renewable sources. Currently, solar electricity and solar hot water projects are eligible for a 30 percent investment tax credit.[25] This credit is available for residential rooftop solar as well as utility-scale solar projects (for example, a solar farm).

The tax subsidy for wind operates differently. A wind project that began construction in 2016 can earn a production tax credit of 2.3 cents per kilowatt-hour of electricity generated during its first 10 years of operation. This is over and above the revenue it gets from selling electricity into the grid.

Subsidies to clean energy are problematic. The first and most obvious problem is that subsidies *lower* the end-user price of energy rather than *raise* it. In Texas, a wind-rich area with much installed wind capacity, generators have willingly accepted a *negative* price for their electricity when demand was very low, say in the middle of the night. This is because the wind generators have next-to-zero operating costs and can collect 2.3 cents in production tax credits for every kilowatt-hour they sell. Even if they have to pay a penny to provide electricity, they are still earning 1.3 cents on each kilowatt-hour sold after cashing in on the production tax credit.[26]

Lowering consumer prices encourages more energy use. It also means that consumers buy fewer energy-efficient appliances and that factory owners invest less in energy-efficient equipment. Subsidies are also expensive. Production and investment tax credits reduce U.S. federal tax collections by about $3 billion a year (Metcalf 2018).

Subsidies also have other problems. They pick winners and losers among competing technologies—thus violating technological neutrality. If the goal is to cut carbon emissions, we should reward technologies that cut emissions regardless of how these technologies work.

Another problem with subsidies is that they are wasteful, with a significant share of the subsidy going to inframarginal purchasers of the capital asset. Consider the $7,500 subsidy for the purchase of a plug-in hybrid vehicle. If the subsidy induces only one in five people to buy a plug-in hybrid, then the effective cost is five times the subsidy, or $37,500—more than the cost of low-end plug-in hybrids.

25. The taxpayer must have adequate tax appetite to use the credit. If tax credits exceed taxes owed, the excess credit can be carried forward and used it in future years. Alternative minimum tax considerations historically also affected the ability to use tax credits, as discussed by Carlson and Metcalf (2008).

26. The problem is not unique to Texas. Wald (2012) reports that the Chicago area experienced negative pricing 3 percent of the time in 2010.

The problem is that we cannot target the subsidy to the prospective car buyer who will be motivated to buy only because of the subsidy. So every buyer gets it. We do not really know whether half the sales would have occurred without the subsidy or if 80 percent of the sales would have occurred without the subsidy. For newer, innovative technologies, one-half may be the right number. But for more common technologies, like energy-efficient windows and appliances that have been subsidized through the tax code, a rule of thumb that four out of five of the sales would have taken place anyway is more reasonable.[27]

Besides being wasteful, energy subsidies disproportionately accrue to high-income households. A 2016 analysis of tax returns shows that 10 percent of energy tax credits go to the bottom 60 percent of the income distribution, while nearly two-thirds go to households in the top 20 percent.[28]

Subsidies can also interact with regulations in unexpected ways. For example, policies that appear complementary can actually undercut each other. Consider the federal tax credit for plug-in hybrids and electric cars. This credit makes it more attractive to buy electric cars and plug-in hybrids. Meanwhile, auto manufacturers are subject to fleet-wide fuel economy standards under the federal CAFE program. For every Chevrolet Volt bought in Massachusetts in part because of the federal credit, General Motors can now sell a gas-guzzling car to someone elsewhere. The purchase of the Volt raises the overall fuel economy of the fleet, and General

27. This may be too conservative. Consider energy-efficient windows. Let us say that a homeowner spends $2,000 to replace older windows with energy-efficient windows. A tax credit (that expired at the end of 2016) worth $200 was available for those windows. Assuming a (generous) price elasticity of –1.0, meaning that demand rises by 1 percent for each 1 percent reduction in price, this credit would induce just over 10 percent in new sales. In other words, 9 sales out of 10 would have occurred in the absence of the subsidy. So, for the one sale of $2,000 in energy-efficient windows that was generated by the tax credit, the government paid out $2,000 in tax credits for windows. This is consistent with the findings in Houde and Aldy (2017), that 70 percent of consumers claiming rebates for an energy-efficient appliance would have bought them anyway, and another 15 to 20 percent simply delayed their purchase by a couple of weeks to become eligible for the rebate. Other research showing a high fraction of purchases that benefit from but are not influenced by a subsidy include studies by Chandra, Gulati, and Kandlikar (2010) and Boomhower and Davis (2014).

28. This study was done by Borenstein and Davis (2016). Some tax credits are more regressive than others. The researchers document that 90 percent of the credits for electric vehicles go to households in the top 20 percent of the income distribution.

Motors is subject to a nationwide mandate on the overall fuel economy of the vehicles it sells.[29]

II.D. Information and Voluntary Programs

Energy experts and policymakers have increasingly focused on the potential for carefully packaged information to reduce energy consumption. Although information is valuable, it is not a viable climate policy. Hunt Allcott and Todd Rogers (2014), for example, show that these programs yield about a 2 percent savings in energy—helpful, but not an approach that is going to get us to a zero-carbon economy.

Offsets are another popular voluntary program. A carbon offset is a payment someone can make to a company to reduce emissions to offset the buyer's own emissions. The problem with offset programs is that it is difficult, if not impossible, to verify that real emission reductions will occur from an offset payment. Moreover, trading in offsets is minuscule relative to the emissions reduction need.[30]

III. Carbon Taxes around the World

Carbon taxes have been used by countries and subnational governments for more than 25 years. As of early 2019, 27 national or subnational carbon taxes were currently in effect or in the process of implementation.[31] There have been two waves of carbon tax enactments, First, a Scandinavian wave starting in the early 1990s saw carbon taxes legislated in Denmark, Finland, Norway, and Sweden, among other countries. By 2000, 7 countries had a carbon tax. A second wave in the mid-2000s saw carbon taxes put in place in Switzerland, Iceland, Ireland, Japan, Mexico, and Portugal. In addition, the Canadian provinces British Columbia and Alberta have

29. It is actually better than that for General Motors. For GHG emissions fleet limits, the EPA treats each 2017 plug-in hybrid sold as if it were 1.7 cars. Electric cars are treated as two cars. And they have a low emission factor (zero for electric), even if the electricity that charges the batteries comes from coal-fired power plants. For fuel economy, the National Highway Transportation Safety Administration, the agency in charge of overseeing fuel economy standards, does not apply a multiplier but does ramp up the fuel economy by dividing the car's estimated fuel economy by 0.15. So an electric car that is rated at 45 miles per gallon gets treated as if it gets 45/0.15 = 300 miles per gallon. For more information, see Center for Climate and Energy Solutions (n.d.).

30. I discuss this in greater detail in Metcalf (2019).

31. Existing and planned carbon tax regimes are summarized by the World Bank Group (2018).

enacted carbon taxes. In 2019, Argentina implemented a carbon tax, and Singapore and South Africa are scheduled to implement carbon taxes in 2019. A South African parliamentary committee moved carbon tax legislation forward so that the full Parliament may consider the tax sometime in 2019 (Szabo 2019). Globally, tax rates range widely, from Poland's carbon tax rate of less than \$1 per ton of CO_2 to as much as \$140 per ton for Sweden. A total of 12 countries have carbon tax rates of at least \$25 per ton, and 6 have rates of at least \$50 per ton.[32]

Given the range in carbon tax rates around the world, how should the United States set the tax rate if it implements a carbon tax? Pigouvian theory suggests the tax on carbon pollution should be set equal to the marginal damage from one more ton of CO_2 emissions.

In a world with preexisting market distortions, economists have argued that the optimal tax on pollution (of any type) will typically be less than the marginal damage.[33] Specifically, the optimal tax equals the marginal damage of pollution divided by the marginal cost of public funds. The larger are the tax distortions, the larger is the marginal cost of public funds and the smaller is the optimal tax relative to marginal damage.[34]

Whether one uses a first- or second-best Pigouvian approach, policymakers need an estimate of the marginal damage from CO_2 emissions. They could base their estimate on analyses of the social cost of carbon done by the EPA and other federal agencies during the Obama administration. This is a measure of damage designed for use in regulatory cost-benefit analyses as opposed to the Pigouvian prescription to measure the social marginal damage of emissions at the optimal level of emissions. The errors in measuring social marginal damage at current emission levels rather than optimal levels are likely to be swamped by errors in estimation from our

32. Rates are as of April 1, 2018, as reported by the World Bank Group (2018).

33. The first papers to make this point were those by Bovenberg and de Mooij (1994) and Parry (1995).

34. See Bovenberg and Goulder (2002) for a review of the literature on second-best environmental taxation and, in particular, section I. As a central case, Bovenberg and Goulder (1996) estimate the marginal cost of public funds to equal 1.25, which suggests that the optimal tax on pollution should be 20 percent lower than social marginal damage. The first-best rule that sets the tax on pollution equal to social marginal damage can be recovered if households have identical tastes, leisure is weakly separable from pollution and private goods, and a nonlinear income tax can be imposed such that the benefits of the pollution tax are exactly offset by the income tax to achieve distributional neutrality. See, for example, Kaplow (1996) and Pirttilä and Tuomala (1997). As Bovenberg and Goulder (2002) point out, these conditions—especially the last—are unlikely to be met.

imperfect state of knowledge about the full range of damage and risks of catastrophic events—events with a high impact but low probability.[35] With this caveat in mind, a tax rate based on the social cost of carbon would be roughly $50 a metric ton of CO_2 in 2020.[36]

A second approach would be to set a tax rate to hit a revenue target over a 10-year budget window. The U.S. Department of the Treasury study projects that a carbon tax starting at $49 a metric ton in 2019 and rising at 2 percent (real) annually would raise $2.2 trillion in net revenue over the 10-year budget window (Horowitz and others 2017). This is net of reductions in other tax collections due to the carbon tax.

Alternatively, a sequence of tax rates could be set over time to achieve a given reduction in emissions by some date. International climate negotiators have focused on a global goal of reducing emissions by 80 percent relative to 2005 by 2050. The United States set this as an aspirational goal in the promises it made in 2015 as part of the international climate negotiations that led to the Paris Agreement. Economic and engineering analyses suggest that an 80 percent reduction by 2050 is possible but would require significant advances in technology along with strong political will.[37] Whether policymakers settle on an 80 percent reduction by 2050 or some other target, a carbon tax will likely be designed with some emissions reduction target in mind.

Let us assume this is the case. How do you ensure you hit the target given our use of a carbon tax? One way to do this is to enact a carbon tax with a "policy thermostat" that adjusts the tax rate in a known and

35. Much has been written on the implications of high-impact, low-probability events—sometimes referred to as fat-tail events. See Wagner and Weitzman (2015) for a lively summary of the literature and a clear statement of the view that climate policy should be seen as an insurance policy rather than as a Pigouvian price adjustment.

36. The $50 figure is based on the estimate by the U.S. Interagency Working Group on the Social Cost of Carbon (2016) for 2020 equal to $42 in 2007 dollars. I have converted the estimate to 2020 dollars using the Consumer Price Index deflator. This is not precisely the right estimate given the methodology used by the Interagency Working Group, but it is close enough given the uncertainties discussed in the text. This also ignores second-best considerations that cause estimates of the optimal tax on emissions to fall short of social marginal damage, as discussed in the notes above. Pindyck (2017) is a prominent critic of using the Interagency Working Group's methodology to set the tax rate on carbon dioxide.

37. Heal (2017) argues that an 80 percent reduction by 2050 could be achieved at "reasonable cost"; he estimates a cost of about 1 percent of GDP. His scenario, however, requires strong financial incentives and political support along with significant reductions in the cost of renewables and battery storage. Williams and others (2014) come to a similar conclusion.

predictable way between now and some future date to increase the likelihood of hitting emission reduction targets 15 to 30 years out.[38]

Next, I describe three carbon tax systems in some detail. They are unique in various ways. British Columbia has a carbon tax on emissions associated with provincial consumption; its tax is one of the most broad-based carbon taxes in place. Switzerland's carbon tax has a unique feature: a tax rate that is adjusted statutorily if emission reduction goals are not met. Sweden's carbon tax has the highest rate in the world, and it has gradually moved to eliminate all discounted rates for energy-intensive sectors subject to the tax.

III.A. British Columbia

As part of a broader package of tax reforms, the Canadian province of British Columbia (BC) enacted a broad-based carbon tax in 2008 starting at $10 (Canadian; hereafter, C$) per metric ton of CO_2 and increasing by C$5 per year to its current C$35 (as of 2018), equivalent to US$27.[39] The tax is scheduled to increase by C$5 per year until it reaches C$50 per ton in 2021. The tax is a broad-based tax on the carbon emissions of all hydrocarbon fuels burned in the province. Given the existing federal and provincial taxes already in place, the carbon tax raised the overall excise tax on gasoline by roughly one-fifth.

The tax collects over C$1 billion annually—over 5 percent of provincial tax collections—and all the revenue is returned to businesses and households through a combination of tax rate reductions, grants to businesses and households, and other business tax breaks (British Columbia Ministry of Finance 2019). Worried that the new carbon tax would disproportionately affect low-income households, policymakers included several elements in the tax reform to offset adverse effects on them. One element was a low-income climate action tax credit of C$154.50 per adult plus C$45.50 per child (as of July 2019), which reduces taxes by C$400 for a low-income family of four. In addition, when first implemented, tax rates in the lowest two tax brackets were reduced by 5 percentage points (Harrison 2013). Also, in the first year of the carbon tax, there was a one-time "climate action dividend" of C$100 for every resident of BC.

38. I propose such a rate adjustment mechanism, called the Emissions Assurance Mechanism, in Metcalf (forthcoming).

39. All currency conversions to U.S. dollars (C$1 = US$0.78) use exchange rates as of late May 2018. Information about the tax rate is taken from https://www2.gov.bc.ca/gov/content/environment/climate-change/planning-and-action/carbon-tax.

This equal-sized dividend represents a greater share of the disposable income of low-income households than that of higher-income households.

Meanwhile, business tax rates were cut. The tax rate for small businesses, for example, was cut from 4.5 percent to 2.5 percent in 2008. As the carbon tax rate rose from C$10 to C$20, there was more carbon tax revenue to rebate, much of which was channeled to businesses in the form of new business tax credits.

BC's carefully constructed policy package to return tax revenue to its residents and businesses balanced concerns about distributional effects and economic growth. Targeting tax cuts to low-income households ensured that the burden of the tax would not fall disproportionately on these households. And the focus on small business emphasized the importance of supporting economic growth.

Canada has moved to a national price on carbon pollution. As of April 2019, every province was required to have a plan in place to price carbon emissions. Failing that, the national government will impose a tax at C$20 per metric ton (Wingrove 2019). Because BC has a carbon tax in place, the federal tax will not be operative in the province.

III.B. Switzerland

Switzerland introduced a carbon tax in 2008 on fuels used for stationary sources (that is, not transportation). Carbon-intensive firms can opt out of the tax in return for committing to specific emission reductions or—for large, energy-intensive firms—by participating in the Swiss cap-and-trade system.[40] One-third of the revenue collected—up to 450 million Swiss francs (hereafter CHF)—is allocated to building efficiency and renewable energy programs. A small amount (CHF 25 million) is set aside for a technology fund. The remainder is redistributed to the public through lump-sum payments to individuals and employer payroll rebates. In 2014, for example, businesses received a payroll rebate of 0.573 percent, while participants in the Swiss mandatory health insurance system received a rebate of CHF 46 per insured person (Carl and Fedor 2016).

In addition to rebating revenue in a lump-sum fashion to businesses and individuals, the Swiss carbon tax is distinctive in linking its tax rate to emission reduction goals. An emissions target provision was added in the 2011 revision of the law: if emissions in 2012 exceeded 79 percent of

40. Information about the Swiss carbon tax comes from the Swiss Federal Office of the Environment at https://www.bafu.admin.ch/bafu/en/home/topics/climate/info-specialists/climate-policy/co2-levy.html.

Table 1. The Swiss Carbon Tax[a]

Tax rate (CHF)	*Enactment date*	*Trigger for a tax rate increase*
12	2008	Not applicable
36	2010	Not applicable
60	2014	Tax rises to CHF 60 if emissions exceed 79 percent of 1990 emissions in 2012
84	2016	Tax rises to CHF 72 if emissions exceed 76 percent of 1990 emissions in 2014
		Tax rises to CHF 84 if emissions exceed 78 percent of 1990 emissions in 2014
96	2018	Tax rises to CHF 96 if emissions exceed 73 percent of 1990 emissions in 2016
		Tax rises to CHF 120 if emissions exceed 78 percent of 1990 emissions in 2016

Sources: International Energy Agency (2018a); Swiss Carbon Tax Ordinance.
a. CHF = Swiss francs. All tax rate changes go into effect at the beginning of the year.

1990 emissions, the tax rate would increase to CHF 60 as of January 1, 2014. Emissions did overshoot the target, and the tax rate was increased. Subsequent tax rate increases in 2016 and 2018 were predicated on emission targets, as detailed in table 1. The current tax rate in 2019 is CHF 96 (US$99).[41] The Swiss tax provides an example of a hybrid carbon tax where rates adjust in response to deviations from desired targets (hence, it is a hybrid of a tax and cap-and-trade system). I discuss a possible hybrid carbon tax design feature in section V below.

III.C. Sweden

Sweden enacted a carbon tax in 1991 as part of a wave of early carbon tax adoptions. Like many other early enactors, it used the revenue to lower marginal income tax rates. The general tax rate rose from a rate of SEK 250 (US$27) to its current rate of SEK 1180 (US$127).[42]

Sectors covered under the EU's ETS are exempt from the tax. Other industrial sectors were initially subject to a lower rate (one-quarter of the standard rate). The rate differential was gradually narrowed, until it was eliminated in 2018.[43] Although the general rate today is 4.72 times its

41. Conseil Federal Suisse, "Ordonnance sur la Reduction des Emissions de CO_2," enacted December 23, 2011 (RS 641.71). Tax rates were reported by the International Energy Agency (2018a, 278). The currency exchange rate is as of mid-September 2018.

42. Exchange rate of SEK 1 = US$0.11, as of February 13, 2019.

43. This information is from https://www.government.se/government-policy/taxes-and-tariffs/swedens-carbon-tax/.

initial rate, carbon tax collections in 2017 were 3.4 times collections in 1994 (the first year for which the Swedish tax authority published data).[44] The slower growth in collections despite the gradual narrowing of the rate differential between the general tax rate and the lower industrial rate reflects reductions in emissions in the Swedish economy.

Sweden is notable for having one of the highest (if not the highest—depending on exchange rate) carbon tax in the world. Its GDP has grown by nearly 80 percent since it enacted a carbon tax in the early 1990s, while its emissions have fallen by one-quarter.[45] Sweden's growth rate has exceeded that of the United States since 2000, despite high taxes on carbon pollution, in part because Sweden uses the revenue to cut other taxes. And the World Economic Forum (2018) finds the two economies to be about equally competitive. The Swedish economist Thomas Sterner notes that though fossil fuels used for home heating are part of the tax base, little in the way of a carbon tax is collected on home heating fuels due to a shift away from fossil fuels for this purpose, a shift that Sterner argues is due largely to the carbon tax.[46]

Runar Brännlund, Tommy Lundgren, and Per-Olov Marklund (2014) find that between 1990 and 2004, Swedish manufacturing output rose by 35 percent while emissions fell by 10 percent, for a 45 percent improvement in emissions intensity. Regression analysis finds that the carbon tax played a significant role in explaining this improvement in emissions intensity. The electric, chemical, and motor vehicle sectors had the highest improvements in emissions intensity, while paper and pulp had the lowest improvements in emissions intensity (albeit a positive improvement).

IV. Economic Outcomes of Carbon Taxes

The literature on the economic effects of carbon taxes is somewhat thin, in part because few broad-based carbon taxes have been in place for a long enough time to assess. Here, I present some regression estimates for emissions and GDP for the Canadian province of British Columbia. Its tax, which has been in place since 2008, is a broad-based assessment on fossil

44. Carbon tax data were downloaded from https://skatteverket.se/omoss/varverksamhet/statistikochhistorik/punktskatter/energiskatterochandramiljorelateradeskatter.4.3152d9ac158968eb8fd24b2.html.

45. The Swedish GDP data are from the World Bank, and the emissions data are from Statistics Sweden (http://www.statistikdatabasen.scb.se).

46. Personal communication, February 12, 2019.

fuels consumed in the province (based on carbon content). I also report evidence from studies of other taxes.

In addition to econometric studies, I report the results of recent modeling economic efforts. The Stanford Energy Modeling Forum (EMF) recently completed a major study (EMF 32) of the economic outcomes of a U.S. carbon tax (Fawcett and others 2018). James McFarland and others (2018) describe the study and the 11 economic models that it analyzed. Results from economic modeling (typically, computable general equilibrium models) are useful, in that they can model technology innovation and general equilibrium responses that econometric studies typically do not. Conversely, model results are driven by model assumptions, which may not always be perfectly transparent.

IV.A. Emissions

Alexander Barron and others (2018) summarize results from Stanford University's EMF 32 study of a U.S. carbon tax. The 11 models participating in the study found that a carbon tax implemented in 2020 at $25 per ton on energy-related fossil fuels would immediately reduce emissions by 6 to 18 percent.[47] A tax of $50 per ton yields a decrease of 11 to 25 percent in emissions in 2020. Over a 10-year period, the models analyzed in the EMF study find that a carbon tax starting at $25 per ton and rising at an annual real rate of 1 percent would lower emissions over the decade (relative to the reference scenario) by 11 to 30 percent, depending on the model, with an average decline of 18 percent. For a carbon tax of $50 per ton rising at 5 percent a year, the 10-year emissions decline ranges from 22 to 38 percent, with an average of 30 percent.

The immediate declines are quite large and likely reflect fuel-switching in the electricity sector as natural gas drives coal out. To appreciate the magnitude of the immediate impact (and the effects over the decade), consider the following calculation. The aggregate consumer price of fossil fuels in 2020, based on the reference scenario of the U.S. Energy Information Administration's (EIA's) (2018) *Annual Energy Outlook*, is $13.87 per million British thermal units (BTUs).[48] Based on the average CO_2 content of each fossil fuel, a carbon tax of $25 ($50) translates

47. Barron and others (2018, 9) report emission reductions of 16 to 28 percent below 2005 levels. Reference-level emissions are about 10 percent below 2005 emissions, according to McFarland and others (2018, figure 2).

48. Prices are consumer prices for nonmetallurgical coal, gasoline, and natural gas (table 3). Consumption shares on a BTU basis are used to average the prices (table 1).

into about \$1.86 (\$3.73) per million BTUs of fossil fuel consumption. A carbon tax of \$25 per ton would increase the consumer price of fossil fuel energy by about 13 percent if fully passed forward to consumers. This suggests an emissions price elasticity of $-.12/.13 \cong -1.0$, using the midpoint of the immediate emission reduction estimates. The 10-year elasticity (based on the average of the study estimates) is about –1.5. Using the carbon tax of \$50 a ton, the immediate emissions price elasticity is about –0.67, and the 10-year elasticity is about –1.11.[49]

Turning to econometric analyses of existing taxes, Boqiang Lin and Xuehui Li (2011) run difference-in-difference regressions of the log difference in emissions in various European countries. Regressions are run for each country individually that imposed carbon taxes in the 1990s—Finland, the Netherlands, Norway, Denmark, and Sweden—with 13 European countries selected as controls. Regressions are run over the 1981–2008 time frame. In 4 of the 5 countries, the growth rate of emissions falls by between 0.5 and 1.7 (based on the estimated coefficient of the interaction variable). Only the estimate for Finland is statistically significant at the 10 percent level, with the coefficient suggesting a drop in the growth rate of emissions of 1.7 percent. The coefficient for Norway is positive but trivially small and statistically insignificant at the 10 percent level. These researchers argue that the larger effect for Finland reflects the smaller number of exemptions from the tax than in other countries.

Ralf Martin, Laurie de Preux, and Ulrich Wagner (2014) consider the impact of the United Kingdom's Climate Change Levy (CCL) on various manufacturing firms' energy and emissions indicators. Adopted in 2001, the CCL is a per-unit tax on fuel consumption by industrial and commercial firms. Unlike a carbon tax, the rate per ton of carbon emissions varies across fuels, from a low of £16 per ton for industrial coal use to a high of £30 (natural gas) and £31 (electricity), as reported by Martin, de Preux,

49. The \$25 carbon tax is modeled to grow at 1 percent real, so it equals \$28 at the end of the decade. The \$50 rate is modeled to grow at a real 5 percent and equals \$81 at the end of the decade. If I compute the 10-year elasticity for the \$50 rate using the average of the initial and final rates, I get a price elasticity estimate of about –0.86. An early study of an actual carbon tax was the study of the Norwegian carbon tax undertaken by Bruvoll and Larsen (2004). They estimate that emissions fell by 2.3 percent relative to a counterfactual of a zero-carbon tax between 1990 and 1999, with changes in the energy mix and energy intensity driving the decline. The Norwegian carbon tax varies across fuels with the 1999 rate, ranging from \$51 a metric ton for gasoline to \$10–19 for heavy fuel oils. Coal for energy purposes was taxed at \$24 a ton. Bruvoll and Larsen estimate an average tax across all sources in 1999 of \$21 a ton. Roughly two-thirds of Norwegian CO_2 emissions were subject to some level of tax.

and Wagner (2014, table 1). They find that CO_2 emissions fall by 8.4 percent, albeit imprecisely estimated. Given the differential carbon tax rates on electricity (£31 per ton) and coal (£16 per ton), we cannot rule out the possibility that the CCL has led to fuel substitution away from electricity and toward coal.[50]

Nicholas Rivers and Brandon Schaufele (2015) consider the impact of BC's carbon tax on the demand for gasoline in the province using data at the province-month level between January 1990 through December 2011. The authors regress log consumption on a carbon-tax-exclusive price of gasoline and a price on the carbon contained in gasoline (based on the tax rate). Although an increase of 1 cent per liter in the price of gasoline depresses gasoline consumption in BC by 0.41 percent, an increase of 1 cent per liter in the carbon tax reduces demand by 1.7 percent—a fourfold increase. The authors attribute the difference to the high salience of the carbon tax.

Looking at province-level emissions, Stewart Elgie and Jessica McClay (2013; updated by Elgie 2014) show that 2013 per capita fuel use subject to the carbon tax declined by over 15 percent relative to 2007 levels, while comparable fuel use in the rest of Canada rose modestly. They did not control for other factors that could affect fuel consumption in Canadian provinces, so it is not clear how much weight to put on these results.

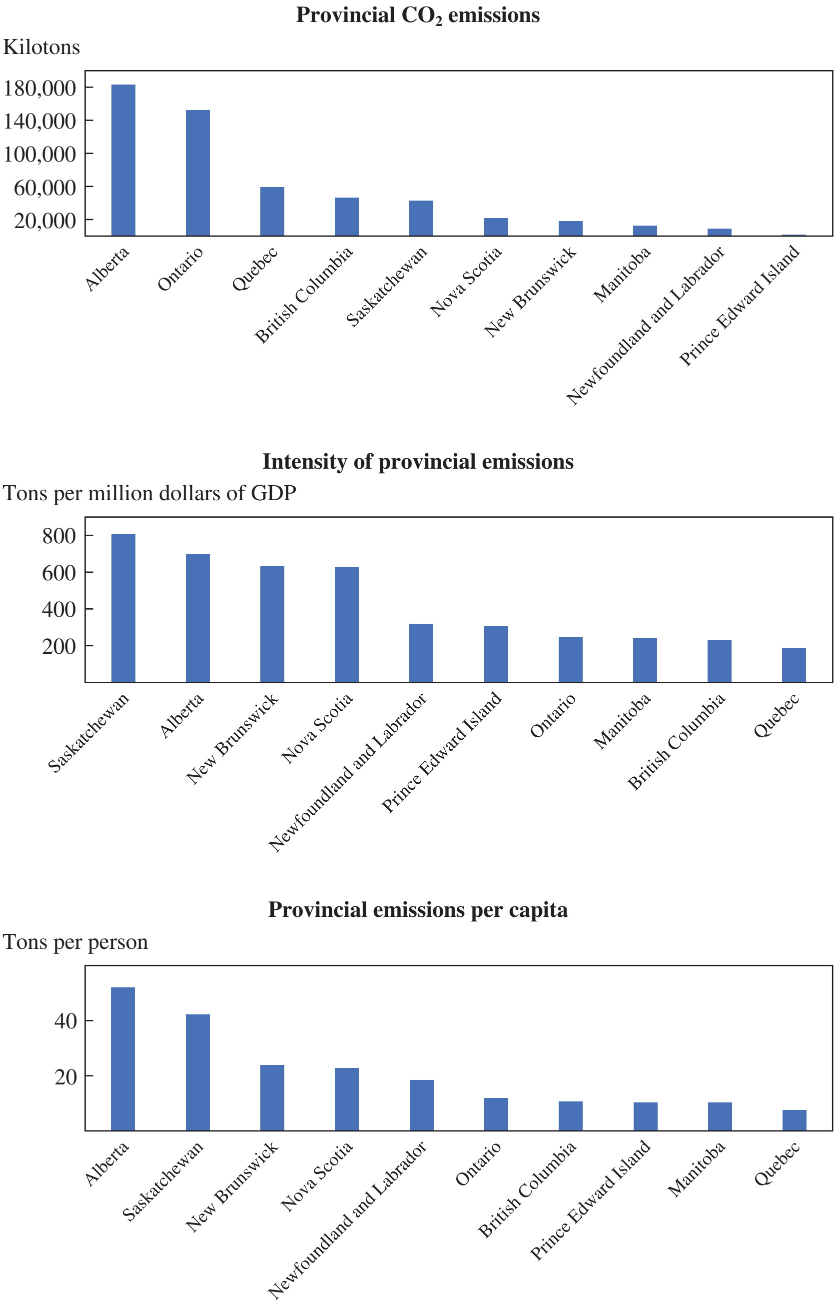
I next present some regressions on annual province-level CO_2 emissions over the period 1990–2016. I present difference-in-difference regressions for a BC carbon tax treatment relative to provinces and territories that have not implemented some form of carbon pricing as well as regressions with carbon prices for the carbon pricing programs in BC, Quebec, and Alberta.

Alberta imposed a price on emissions in July 2007 called the Specified Gas Emitters Regulation. In effect, it is a carbon-intensity cap-and-trade program (Leach 2012). Quebec implemented a modest cap-and-trade program in 2013.

Before running regressions, it is worth noting that though BC was a moderately large source of CO_2 emissions in Canada in 2007 (the top panel of figure 2), it is a small emitter on a per capita basis (the middle panel of figure 2) or per dollar of GDP (the bottom panel of figure 2). It is perhaps not surprising that three of the four provinces that have moved forward with carbon-pricing programs (BC in 2008, Quebec in 2013, and Ontario in 2017) have very low emissions per capita or low emissions intensity. Alberta, conversely, is a top emitter on nearly all three metrics.

50. The coefficient on the treatment variable in a regression with a measure of solid fuel use (coal and coke) as the dependent variable is positive but not statistically significant.

Figure 2. Provincial Measures of CO_2 Emissions in Canada, 2007



Source: Appendix, table A2.

Table 2 presents CO_2 emission regressions for the Canadian provinces and territories over the period 1990–2016.[51] I include a treatment dummy for the BC carbon tax as well as controls for GDP, population, and trade. For the latter, I include an export index variable that measures the price of goods exported from each province, weighted by province-level exports. All regressions include province and year fixed effects.

The first regression includes all provinces and territories and finds a treatment effect of –3.6 percent, albeit imprecisely estimated. This is likely to be biased upward as I am including provinces in the control group that have put a price on carbon. In column 2, I exclude Alberta, Quebec, and Ontario. The first two provinces put a price on emissions during the control period. Ontario is excluded because it has an ambitious feed-in tariff for renewable energy (enacted in 2009) that is unique among Canadian provinces.[52] Dropping these three provinces increases the impact of the BC carbon tax. Now emissions fall in the posttax period by 6.6 percent. If I limit the regression period to 1995–2016, the impact is even larger (column 3). Columns 4 and 5 run the regression on the log of emissions per dollar of GDP (emissions intensity). With the sample restricted to 1995–2016, the impact is precisely estimated at the 1 percent level.

Table 3 provides results when the carbon prices for Alberta, Quebec, and BC are included.[53] The coefficient on the tax rate variable is consistently negative across the regressions but only statistically significant when the time frame is limited to 1995–2016. Focusing on the coefficient in column 2, a \$30 carbon tax (BC's rate in 2012) reduces emissions by 7.8 percent, a result consistent with the results in table 2.

Although the regression results given in tables 2 and 3 are not precisely estimated across the board, they tell a consistent story of the tax reducing emissions in BC of between 5 and 8 percent since the tax went into effect in 2008.

IV.B. GDP

Table 4 reports similar regressions with ln(GDP) as the dependent variable. Unlike the emission regressions, I also consider variables that

51. The data sources for the regressions in tables 2 through 5 are given in the appendix at the end of this paper, in table A2.

52. Ontario's feed-in tariff is described at https://www.ontario.ca/document/renewable-energy-development-ontario-guide-municipalities/40-feed-tariff-program.

53. Quebec's rate is C\$3.50 starting in 2007. A cap-and-trade system went into effect in 2013, and I include average allowance auction prices for each year. Alberta enacted the Specified Gas Emitters Regulation in 2007 at a rate of \$15 per ton.

Table 2. Carbon Dioxide Emission Regressions: British Columbia (BC), Difference-in-Difference[a]

	(1)	*(2)*	*(3)*	*(4)*	*(5)*
BC treatment	–0.036	–0.066*	–0.088***	–0.057*	–0.073***
	(0.024)	(0.036)	(0.026)	(0.027)	(0.022)
GDP	0.624***	0.565***	0.419**	—	
	(0.147)	(0.151)	(0.173)		
Population	0.275	0.491	1.114*	0.178	0.420
	(0.164)	(0.316)	(0.586)	(0.221)	(0.420)
Export price	0.001*	0.002	0.001	0.002*	–0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	–1.089	–3.317	–9.779	3.608	–6.651
	(2.351)	(4.281)	(6.901)	(2.871)	(5.398)
Provinces and territories	All	Excludes AL, ON, QC	Excludes AL, ON, QC	Excludes AL, ON, QC	Excludes AL, ON, QC
Years	1990–2016	1990–2016	1995–2016	1990–2016	1995–2016
Observations	360	279	234	279	234
R^2	0.998	0.996	0.996	0.939	0.9981

Source: Appendix, table A2.

a. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. AL = Alberta; ON = Ontario; QC = Quebec. All regressions include province and year fixed effects. The dependent variable is ln(CO_2 emissions). GDP and population are in logs. Regressions 4 and 5 have the ln of CO_2 per dollar of GDP as the dependent variable. Standard errors are in parentheses and are clustered at the province level.

Table 3. Carbon Dioxide Emission Regressions—Tax Rates[a]

	(1)	*(2)*	*(3)*	*(4)*	*(5)*
Carbon tax rate	–0.0013	–0.0026	–0.0035*	–0.0022	–0.0028**
	(0.0021)	(0.0018)	(0.0017)	(0.0014)	(0.0013)
GDP	0.6230***	0.5697***	0.4536***	—	—
	(0.1465)	(0.1354)	(0.1482)		
Population	0.3017	0.4388**	0.8490**	0.1307	0.3540
	(0.1972)	(0.1816)	(0.3213)	(0.1129)	(0.2173)
Export price	0.0014*	0.0014*	0.0009	0.0017*	0.0014
	(0.0007)	(0.0007)	(0.0008)	(0.0008)	(0.0009)
Constant	–1.4332	–2.6635	–6.8224	–3.0303*	–5.9313*
	(2.5907)	(2.3637)	(3.9806)	(1.4842)	(2.8518)
Provinces and territories	All	Excludes ON	Excludes ON	Excludes ON	Exclude ON
Years	1990–2016	1990–2016	1995–2016	1990–2016	1995–2016
Observations	360	333	278	333	278
R^2	0.998	0.997	0.998	0.958	0.957

Source: Appendix, table A2.

a. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. ON = Ontario. All regressions include province and year fixed effects. The dependent variable is ln(CO_2 emissions). GDP and population are in logs. Regressions 4 and 5 have the ln of CO_2 per dollar of GDP as the dependent variable. Standard errors are in parentheses and are clustered at the province level.

Table 4. GDP Regressions: BC Difference-in-Difference[a]

	(1)	*(2)*	*(3)*	*(4)*
BC treatment	–0.0022	0.0416**	0.0923*	0.0788
	(0.0179)	(0.0144)	(0.0431)	(0.0447)
Canadian GDP	0.8422***	0.8541***	0.8969***	0.8844***
	(0.1044)	(0.0859)	(0.0813)	(0.1426)
Population	0.6153**	0.3987*	0.0615	0.1089
	(0.2645)	(0.2169)	(0.3094)	(0.5356)
Export price	–0.0007	0.0009	0.0010	0.0011
	(0.0007)	(0.0008)	(0.0009)	(0.0009)
Manufacturing share		0.2974	0.2869	0.1756
		(0.3736)	(0.6240)	(0.6226)
Professional share		–1.4859	–2.5594	–2.7270
		(1.0505)	(1.4941)	(1.6554)
Public sector share		–0.7057	–0.0253	–1.1626
		(0.8856)	(0.9117)	(0.8190)
Natural resources share		0.9055	0.1708	0.0537
		(1.5229)	(1.2507)	(1.4702)
Constant	–9.8283***	–6.7350**	–3.1841	–3.5480
	(2.3458)	(2.3089)	(3.4390)	(5.4709)
Provinces and territories	All	Provinces only	Provinces less AL, QC, ON	Provinces less AL, QC, ON
Years	1990–2016	1990–2016	1990–2016	1995–2016
Observations	360	270	189	154
R^2	0.999	0.999	0.999	0.999

Source: Appendix, table A2.

a. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. AL = Alberta; ON = Ontario; QC = Quebec. All regressions include province fixed effects. The dependent variable is ln(GDP). Canadian GDP and population are in logs. Standard errors are in parentheses and are clustered at the province level.

measure the composition of economic activity in provinces and territories. Specifically, I include the share of workers in the employment categories of manufacturing, professional services, the public sector, and natural resources.[54] Regressions include province fixed effects. Rather than year fixed effects, I include Canadian GDP (in logs) to control for business cycle effects at the national level. Column 1 of the table does not include the economy composition variables, and the estimated coefficient on the carbon tax treatment variable is negative, though economically small (–0.22 percent) and imprecisely estimated. The coefficient turns positive and is both economically and statistically significant when the composition variables are

54. Natural resources includes forestry, fishing, mining, quarrying, oil, and gas. I do not include these share variables in the emission regressions, because I would expect the carbon tax to reduce emissions, in part, by shifting the composition of economic activity.

Table 5. GDP Regressions—Tax Rates[a]

	(1)	*(2)*	*(3)*	*(4)*
Carbon tax rate	–0.0005	0.0018*	0.0024	0.0022
	(0.0015)	(0.0009)	(0.0014)	(0.0013)
Canadian GDP	0.8406***	0.8625***	0.8540***	0.8802***
	(0.1067)	(0.0847)	(0.0835)	(0.1099)
Population	0.6294*	0.3600	0.3167	0.3185
	(0.2920)	(0.2246)	(0.2516)	(0.2970)
Export price	–0.0006	0.0009	0.0011	0.0010
	(0.0007)	(0.0008)	(0.0009)	(0.0009)
Manufacturing share		0.3312	0.4136	0.3205
		(0.3599)	(0.4239)	(0.3648)
Professional share		–1.6006	–1.7846	–2.3823*
		(1.0612)	(1.1229)	(1.2465)
Public sector share		–0.6915	–0.6474	–1.1353
		(0.8879)	(0.9814)	(0.6643)
Natural resources share		0.7830	0.6903	0.4506
		(1.4176)	(1.4094)	(1.3736)
Constant	–9.9960***	–6.2940**	–5.7433*	–6.0458*
	(2.6875)	(2.4443)	(2.8014)	(2.8513)
Provinces and territories	All	Provinces only	Provinces less ON	Provinces less ON
Years	1990–2016	1990–2016	1990–2016	1995–2016
Observations	360	270	243	220
R^2	0.999	0.999	0.999	0.999

Source: Appendix, table A2.

a. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. ON = Ontario. All regressions include province fixed effects. The dependent variable is ln(GDP). Canadian GDP and population are in logs. Standard errors are in parentheses and are clustered at the province level.

included. Columns 3 and 4 exclude Alberta, Quebec, and Ontario. When regressions are run over the 1990–2016 period, the estimated change in GDP is 9.23 percent and is significant at the 10 percent level. When the regression is limited to 1995–2016, the coefficient falls to 7.88 percent and just misses being statistically significant at the 10 percent level.

Table 5 repeats regressions with the carbon tax rate for all provinces with carbon pricing in effect. The coefficients on the tax rate are not statistically significant but tell a similar story as in table 4. A $30 carbon tax is associated with a roughly 6 percent increase in GDP.[55] These GDP results are consistent with simpler regressions run in my 2016 paper, although

55. These regressions suggest that the BC carbon tax led to higher GDP. Regressions not reported here suggest that the tax may have raised the growth rate of BC's GDP by as much as 1 percent.

those results were an order of magnitude smaller. Given the imprecise estimates, we should not lean too heavily on these results. But it seems fair to say that GDP has not been adversely affected by the carbon tax. A couple of factors about the BC carbon tax support this result. First, the tax was designed to be revenue neutral, with some of the revenue used to lower personal and business tax rates. This should enhance the efficiency of the provincial economy and could have a positive impact on growth. Second, some of the revenue was specifically directed to lower-income households. To the extent that these households have higher marginal propensities to consume out of income, this could, as well, support economic growth in the short run.

As additional evidence on the GDP effects of a carbon tax, I provide analysis using variation in carbon tax implementation in European countries; see table 6 for the regression results.[56] I focus on countries that are part of the ETS, a cap-and-trade system covering the power sector and certain other energy-intensive sectors (see above).[57] These countries have a uniform treatment of emissions under the cap-and-trade system. Fifteen of these countries have enacted carbon taxes on top of the ETS, covering sectors or firms within sectors not covered by the ETS. Although one should be cautious in interpreting results of regressions of GDP on an indicator for the presence of a carbon tax as causal, the regressions can shed light on whether GDP is adversely affected by the presence of a carbon tax. Data on 31 countries are analyzed over the period 1985–2017. The first carbon tax in the sample went into effect in 1991.

The first regression shown in table 6 regresses the log of real GDP against an indicator variable for the presence of a carbon tax. The regression includes Organization for Economic Cooperation and Development (OECD)–wide ln(GDP) and country fixed effects. The GDP effect is positive, with a 3.89 percent increase in EU country GDP, but is not statistically significant. The second regression adds a variable interacting the indicator with a variable measuring the share of the country's emissions covered by the carbon tax at the beginning of 2019.[58] In contrast to the BC carbon tax, which applies to all emissions in the province, carbon taxes vary across Europe in scope of coverage. To capture differential coverage, I include

56. Data sources for these regressions are given in the appendix, in table A3.

57. I also include Switzerland, which has its own cap-and-trade system that is closely aligned with the ETS. The two systems will be formally linked starting in 2020.

58. The World Bank's Carbon Pricing Dashboard maintains information on current carbon tax rates and coverage. Its data go back to 2016. Data on earlier years are not available.

Table 6. GDP Regressions for the European Union[a]

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Carbon tax indicator	0.0389	0.0814	0.0335	0.0309	0.166	0.106
	(0.0545)	(0.137)	(0.131)	(0.0693)	(0.177)	(0.171)
Interaction with emissions share		–0.144	0.0260		–0.447	–0.260
		(0.411)	(0.388)		(0.483)	(0.462)
OECD ln(GDP)	1.178***	1.179***				
	(0.0913)	(0.0907)				
OECD ln(GDP per capita)				1.416***	1.421***	
				(0.119)	(0.119)	
Constant	–8.363***	–8.387***	12.67***	–4.493***	–4.554***	10.65***
	(1.596)	(1.585)	(0.0401)	(1.244)	(1.246)	(0.0435)
Tax effect at median emissions share		0.0341	0.0420		0.0181	0.0203
		(0.0539)	(0.0514)		(0.0607)	(0.0657)
Observations	918	918	918	912	912	912
R^2	0.830	0.830	0.848	0.746	0.751	0.779
Year fixed effects	No	No	Yes	No	No	Yes

Source: Appendix, table A3.

a. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. OECD = Organization for Economic Cooperation and Development. All regressions include country fixed effects. Dependent variable is ln(GDP) in regressions 1 through 3 and ln(GDP per capita) in regressions 4 through 6. Regressions are run for 1985–2017. Standard errors are in parentheses and are clustered at the country level.

this interaction variable. The coefficient on the carbon tax indicator variable is positive, and the interaction coefficient negative. The interquartile range of GDP effects, given the distribution of the share variable conditional on having a carbon tax, runs from 2.4 percent (for the 75th percentile of the share of covered emissions) to 6.0 percent (for the 25th percentile). The impact for the median covered emissions share is 3.4 percent (reported in table 6). In no case is the impact statistically significant at any reasonable level. The third column adds year dummies with no appreciable impact on the effects.

The final three columns of table 6 run regressions on the log of per capita real GDP. The results are not materially different. The regressions, as a group, suggest that imposing a carbon tax has not adversely affected GDP in countries that have levied a carbon tax. If anything, there appears to have been a modest positive impact—if we take the coefficient estimates at face value. I have not explored the mechanism underlying this positive impact (if, indeed, it holds up). Many early carbon tax reforms used carbon tax revenues to lower income tax rates as part of a green tax reform movement in the early 1990s, especially in those Nordic countries with very high income tax rates (Brännlund and Gren 1999). Lowering especially high income tax rates through a carbon tax reform could stimulate economic activity. More ex post analysis of existing carbon tax systems would be extremely valuable, both for assessing the macroeconomic effects of a carbon tax and for calibrating economic models that are typically used to assess climate policy. Such analyses would also be valuable for teasing out the mechanisms driving economic growth—if they hold up in subsequent research.

IV.C. Employment

As part of their analysis of the United Kingdom's CCL, Martin, de Preux, and Wagner (2014) found that the climate levy was associated with an increase in employment, though imprecisely estimated. They conclude that a factor substitution effect (labor for energy) was driving the employment increase in U.K. manufacturing.

Akio Yamazaki (2017) constructs employment data on 68 industries across Canadian provinces and territories for the years 2001–13 to investigate the BC carbon tax's impact on employment. Yamazaki notes that the carbon tax could affect employment by driving up costs and discouraging production and hence employment (output effect). The tax redistribution deriving from how carbon tax revenues are returned to businesses and households could stimulate demand for products and hence workers (a redistribution effect). Finally, employment could rise (or fall) if labor is a substitute (or a complement) for energy (factor substitution effect). His study focuses on the first two channels of employment effects. He finds that the output effect dampens employment while the redistribution effect enhances employment. In the aggregate, he finds a modest positive and statistically significant impact on employment, on the order of 0.75 percent annually. Jobs are shifting, however, from carbon- and trade-sensitive sectors to sectors that are less carbon and trade sensitive. Chemical manufacturing, for example, has the largest decline in employment, while health care has the largest increase.

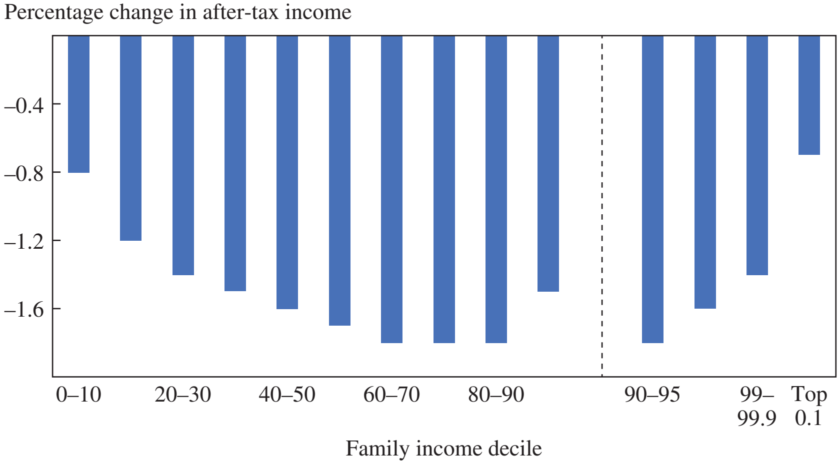
IV.D. Distributional Outcomes

Numerous distributional analyses have been done of a carbon tax for the United States. Distributional effects arise from differential consumption of carbon-intensive goods whose prices have gone up relative to the general price index versus carbon-light goods whose prices have fallen relative to the general price index. This is the *use side* impact, and numerous studies have shown that this distributional channel is regressive. The tax also can lower factor prices. If returns to capital fall more than wages, then the carbon tax will have a progressive aspect on the *sources side*. Another factor contributing to progressivity on the sources side is the existence of indexed transfers that are disproportionately important for lower-income households.[59] Lawrence Goulder and others (forthcoming) show in a computable general equilibrium analysis that the source side effects fully offset the use side effects, so that the carbon tax, ignoring the use of revenue, is distributionally neutral to slightly progressive.

Metcalf (1999), among others, has argued that one should focus on the distributional effects of carbon tax reform, by which I mean the package of a carbon tax and the use of the proceeds, whether it be new spending, tax cuts, or cash grants to households. Distribution of the carbon revenue through an equal per capita cash grant—as proposed by, for example, the Climate Leadership Council—would be highly progressive. Distributional tables from a recent U.S. Treasury research paper (Horowitz and others 2017) illustrate this. Figure 3 shows the carbon tax, ignoring the use of revenue. The Treasury's analysis finds it is progressive up through the

59. Rausch, Metcalf, and Reilly (2011) and Goulder and others (forthcoming), among others, have argued that use-side, regressive effects are offset by progressive, source-side effects. Transfers are also important in explaining the source-side, progressive effects. Fullerton, Heutel, and Metcalf (2011) also stress the importance of transfers.

Figure 3. The Carbon Tax Burden, Ignoring the Use of Revenue



Source: U.S. Department of the Treasury (2017).

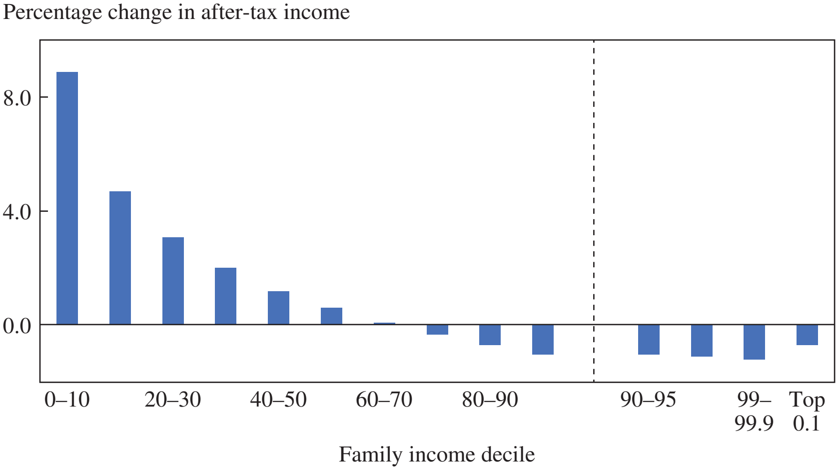
7th and 8th deciles. It then turns regressive in the top deciles. With the equal per capita rebate, shown in figure 4, the tax reform is sharply progressive. In fact, households up through the 70th percentile are better off, in the sense of receiving more in the rebate than the effects on disposable income through source and use side effects. Note, however, that these graphs are showing average distributional effects at each decile. Various researchers have noted that there can be considerable heterogeneity within a decile (Rausch, Metcalf, and Reilly 2011; Cronin, Fullerton, and Sexton 2017).

V. Policy Thoughts

In this paper, I do not address the details of how one would implement a carbon tax. This topic has been covered elsewhere—by Metcalf and David Weisbach (2009), Metcalf (2017), and Horowitz and others (2017). In brief, an excise tax on coal, natural gas, and petroleum products can piggyback on existing fuel excise taxes (for petroleum and coal). Additional process emissions can easily be taxed, such that roughly 90 percent of domestic GHG emissions (excluding forestry and land-use changes) can be included in the tax base.[60]

60. See Metcalf and Weisbach (2009) for further discussion.

Figure 4. The Carbon Tax with Equal Rebates per Person



Source: U.S. Department of the Treasury (2017).

Two design points are worth mentioning. First, any emissions captured and permanently stored should not be taxed. Depending on the locus of taxation, these emissions can either be excluded from the tax base or a rebate of the tax paid at a previous stage of production can be provided to anyone engaging in approved capture-and-sequestration techniques.

Second, a federal carbon tax will need to consider whether and how to tax imported emissions (and how to treat exports of carbon-intensive goods). Ideally, we would tax the carbon content of all imports and exempt from taxation the carbon content of all exports. Doing so would tax emissions associated with domestic consumption. Taxing fossil fuel imports (and rebating the tax on exports) is straightforward and should be part of the tax design. Taxing the embedded CO_2 in imported goods and services is more difficult. Wayne Gray and Metcalf (2017) document that roughly 95 percent of the value of manufacturing shipments has very low carbon content. We need only concern ourselves with a handful of carbon-intensive intermediate and final goods. Determining the carbon content of selected imports is a nontrivial task, and Metcalf and Weisbach (2009) propose setting the tax on the basis of the emissions content of domestically produced carbon-intensive goods.

A carbon tax addresses the central problem of climate change: that the social cost of burning fossil fuels exceeds the private, market cost. A tax is the most flexible way to persuade millions of economic agents to adjust their behavior in large and small ways to reduce emissions. Although pricing our carbon pollution is a necessary element in a cost-effective climate policy, it is not a sufficient policy, for a number of reasons. Other market failures, the existence of GHG pollutants not amenable to taxation, and institutional barriers suggest the need for a range of policies.

As discussed in the introduction, the United States' transition to a zero-carbon economy will require new inventions and production processes. Research and development will be key to the successful diffusion of these technologies. Information and new knowledge are pure public goods that are underprovided in a market economy. A carbon tax should be complemented with a major increase in zero-carbon energy research to help develop cost-effective replacements for fossil fuels.

In addition, various regulatory and other institutional barriers impede the transition to a zero-carbon economy. Resistance by states to interstate transmission lines passing through their state can limit the use of zero-carbon electricity (for example, wind from the Midwest and hydropower from Canada).[61] The lack of clear legal and financial liability rules for carbon capture and sequestration will also impede the growth of this technology when and if it becomes cost-competitive.[62]

Although these other issues are important, putting a price on carbon pollution is central to any effective national policy. How do we overcome the political hurdles and get a carbon tax enacted? It will require strong political leadership. It may be that a framework for reform can also help. A powerful disciplining device for the Tax Reform Act of 1986 was the clear set of guidelines laid out by Ronald Reagan in his 1984 State of the Union Address, where he called for a tax reform that simultaneously lowered tax rates while maintaining revenue neutrality. A similar set of guidelines—or a policy framework—would be useful for carbon tax reform. My policy framework for a national carbon tax includes (1) revenue neutrality, (2) a focus on fairness, (3) streamlined policy, and (4) significant emission reductions.

Revenue neutrality ensures that long-contentious partisan differences over the size of the federal budget should not be allowed to affect the

61. Joskow and Tirole (2005) point out other barriers and market failures that lead to suboptimal investment in transmission lines.

62. The National Academy of Sciences (2019) lays out a research agenda to address the various barriers and high costs of carbon capture and storage.

climate policy debate. A revenue-neutral carbon tax reform disentangles these two issues and may ensure greater bipartisan support for a carbon tax.

Because energy makes up a more significant share of the budget of low-income families than higher-income families, many worry about a carbon tax's impact on poorer households. Tax reform packages can be designed to offset any regressive impact on lower-income households. One could take the approach of the Climate Leadership Council's tax-and-dividend approach and rebate all the revenue to U.S. families. This would have bipartisan appeal. But a carbon tax plan can achieve fairness without necessarily giving all the revenue back through a dividend program. A portion of the revenue could go to low- and moderate-income households to offset higher energy bills, while the remainder could be used to lower income tax rates. Lowering tax rates would disproportionately benefit higher-income households and so ensure benefits across the entire income distribution. Using revenue to lower tax rates would also increase the efficiency of the U.S. economy by reducing disincentives to work or save.

There is another aspect to fairness. How should we treat workers in industries that are disproportionately affected by the shift to a zero-carbon economy? Nearly one-quarter of all U.S. coal miners work in West Virginia. Kentucky, Wyoming, and Pennsylvania together account for one-third of coal-mining jobs. No other state comes close to the number of coal miners in these states. If we focus on a state's dependence on coal rather than on the absolute number of jobs, West Virginia and Wyoming stand out. They have the highest share of employees working in coal mining (2 percent), and diversifying each state's economy to become less dependent on coal would benefit the economies of these states. A national carbon tax proposal should also consider how economic development programs could help coal-dependent regions transition to a postcoal economy.[63]

A carbon tax allows us to eliminate many energy-related tax breaks, starting with tax preferences for oil and gas production in the United States. These cost roughly $4 billion a year (Metcalf 2018) and run counter to good environmental and climate policy. Next, we can remove various investment and production tax credits for renewable energy projects. These tax preferences only make sense to support renewable energy investment

63. All employment data are for 2017. Coal-mining employment is taken from the EIA's *Annual Coal Report 2017*, and total employment is taken from the U.S. Census Bureau's Quarterly Workforce Indicators, which are available at https://qwiexplorer.ces.census.gov.

and production if we cannot tax carbon pollution. The existing tax breaks are a way to level the playing field between carbon-polluting fuels and carbon-free fuels. If we cannot raise the cost of the polluting fuel, then the next best thing is to lower the cost of the nonpolluting fuel. But if we enact a carbon tax, a reasonable bargain is to eliminate those tax preferences, for a savings of roughly $6 billion a year.[64]

Next, consider the Clean Air Act and the endangerment finding that CO_2 should be regulated under the act. Although the idea of replacing an inefficient regulatory approach with an efficient pricing mechanism is appealing, the Clean Air Act has been a powerful tool for improving environmental quality in this country over the past half century. Simply giving up Clean Air Act oversight of carbon pollution is asking quite a bit, given the potential for Congress to pass a carbon tax today only to have a future Congress repeal the tax. The challenge is to construct a carbon tax that provides the assurances that we will meet environmental goals over the course of this century.

One way forward is to preserve the EPA's regulatory authority over GHG emissions but suspend any regulatory action for emissions covered by a carbon tax as long as demonstrable progress in reducing emissions is being made. This, of course, requires that we define "progress." Progress could be measured as a target reduction in emissions relative to a given base year (for example, 2005 emissions) at various milestone years between now and 2050. Failure to hit the targeted emission reductions would automatically trigger resumption of the EPA's regulatory process under the Clean Air Act. An independent commission or advisory group established under law could oversee progress toward the emission reductions. In addition, the carbon tax could be designed so the tax rate automatically adjusts over time to keep the United States on target to reach long-run emission reduction goals.[65]

This is not to argue that *all* GHG regulations should be put on hold. It is not realistic to subject all GHG emissions to a carbon tax. Some emissions are simply too hard to measure. A good example is the methane emissions associated with fossil fuel extraction. Methane is a potent GHG with a short-run impact on the environment 30 times that of CO_2. When underground coal mining was the dominant source of coal in the

64. This is a 10-year average (over the period 2019–28) of the tax expenditure estimates for energy production and investment tax credits, as reported by OMB (2019).

65. Hafstead, Metcalf, and Williams (2017) and Metcalf (forthcoming) lay out the idea of a self-adjusting carbon tax to hit emission targets.

United States, coalbed methane was a major source of GHG emissions. Now, with the shift to surface coal, methane emissions are more associated with oil and natural gas fracking. These emissions are hard to measure and are found at nearly every drilling site to some extent. Rather than try to measure and tax these emissions, it makes more sense to put strong regulations in place that require state-of-the-art drilling and extraction techniques and that equipment be used to minimize methane leaks. This would be coupled with strong monitoring and enforcement. Similarly, agricultural and land use emissions are difficult to tax and thus are more suitable for regulation.

In summary, we need to avoid a "bait and switch" situation, whereby regulatory oversight over GHGs is traded for a carbon tax, only to find that Congress does not have the will to set a sufficiently high tax to make a significant dent in emissions. Many environmentalists are already mistrustful of a carbon tax, and it will be important to bring them on board in order to get Congress to act. This leads to my last framework principle. The policy must significantly cut emissions.

It will not do to set a carbon tax at $25 a ton and simply let it rise at the rate of inflation over time. It is impossible to say exactly what tax rate is required to achieve a particular emissions target. Much depends on technological advancement and consumer behavior. However technology advances, it is likely that we will need a robust carbon price. The 2014 Stanford EMF modeling exercise found that a 50 percent reduction in U.S. emissions by 2050 would require a carbon price between $10 and $60 per ton of CO_2 in 2020 (looking across the bulk of models and technology assumptions) and between $100 and $300 in 2050. Although the international climate negotiations have focused on a target of an 80 percent emissions reduction by 2050 from 2005 levels, most research suggests that this will be extremely expensive. The Stanford modeling study corroborates this. The participating modelers estimate that the 2050 price on CO_2 required to hit that target would be somewhere in the range from $200 to more than $500 a ton, depending on model assumptions.[66]

What carbon price will be needed to reach any future emissions target will depend in large measure on the pace of clean energy technological development. A substantial price on CO_2 emissions will help spur this development. Given the very high (and probably politically unacceptable) cost of an 80 percent emissions reduction, a more modest but still aggressive

66. The Stanford Energy Modeling Forum exercise (EMF 24) is described by Clarke and others (2014).

goal of emission reductions between now and 2050 may be advisable. One approach would be to set a target for 2035 combined with an assessment beginning in 2030 to set a subsequent target for 2050. A 2035 target of a 45 percent reduction in CO_2 emissions (relative to 2005 levels), for example, would be ambitious but within reach. A subsequent target could be set for 2050, with an emissions reduction perhaps somewhere in the range of 60 to 80 percent by 2050, with the precise target set as new information emerges over the first 15 years about the damage from both GHG emissions and clean energy technology costs.[67]

Any target set out in carbon tax legislation could be conditioned on OECD member countries also committing to this goal within a short time frame and the major non-OECD emitting countries committing to this goal within, say, a decade. This could be combined with the Nordhaus (2015) "climate club" idea. Developed countries (or any group of major countries, for that matter) could band together and impose trade sanctions on countries that do not take effective action to reduce emissions.[68]

Once the goal is set, the carbon tax should contain a mechanism for adjustment to ensure that the target is met. One simple way to do that would be to enact a carbon tax with an initial tax rate (for example, $40 a ton of CO_2 emissions). The legislation would also include a clear and transparent rule for adjusting the tax rate over time to hit emission reduction benchmarks, as also set out in the legislation. This would provide greater assurance that the United States would hit desired emission reduction targets while still providing the price predictability that the business community needs.[69]

The carbon tax should also be designed so that there is the political will to sustain high tax rates on emissions. The authors of the Climate Leadership Council's carbon tax and dividend plan argue that the dividend

67. Metcalf (forthcoming) discusses the use of sequential targets for a carbon tax and proposes a 45 percent reduction by 2035 that would be consistent with a 60 percent reduction target by 2050. If clean energy technology costs fall more rapidly than expected, the 2050 target could be strengthened when set in the mid-2030s.

68. Nordhaus argues that nonparticipating countries could be punished with carbon tariffs or a uniform tariff on all imported goods to club members. He finds that a modest uniform tariff is more effective at promoting club membership than a carbon tariff. How Nordhaus's club idea would dovetail with the existing international trade order overseen by the World Trade Organization is unclear.

69. This rate adjustment mechanism is set out in a proposal in Metcalf (forthcoming). His proposal builds on work by Hafstead, Metcalf, and Williams (2017). Other approaches to ensuring greater certainty of given emissions reduction targets are proposed by Aldy (2017; forthcoming) and Murray, Pizer, and Reichert (2017).

will help build political support for high tax rates because, as rates rise, so would dividends.[70] They may or may not be right; but they are focusing on the right question: how to build political will for the changes to our energy system necessary to move to a zero-emissions economy.

VI. Conclusion

A carbon tax is a cost-effective policy tool to reduce the United States' GHG emissions. It would be easy to implement, easy to administer, and straightforward for firms' compliance. With 23 carbon taxes in place around the world, a carbon tax is moving from a theoretical fancy of economists to a political reality. The politics around enacting a carbon tax continue to be challenging, but it is encouraging that bipartisan support for a carbon tax is growing. Although a carbon tax will entail costs to the economy—after all, we cannot clean up the environment for free—evidence from other countries indicates that a carbon tax need not impose large costs on an economy. The evidence from British Columbia suggests, in fact, that a well-designed carbon tax can actually boost jobs and GDP while reducing carbon emissions.

ACKNOWLEDGMENTS James Stock, Larry Goulder, and Adele Morris provided helpful comments on early drafts of this paper. Ozgur Bozcaga provided invaluable research assistance on this project.

70. Baker and others (2017, 3) write: "It is essential that the one-to-one relationship between carbon tax revenue and dividends be maintained as the plan's longevity, popularity, and transparency all hinge on this. Allocating carbon tax proceeds to other purposes would undermine popular support for a gradually rising carbon tax and the broader rationale for far-reaching regulatory reductions."

Appendix

Table A1. Carbon Taxes around the World[a]

Jurisdiction	*Type*	*Year of implementation*	*Price (dollars)*	*Share of jurisdiction's GHG emissions covered*	*Revenue, 2018 (millions of dollars)*
Finland	National	1990	76.87	36%	1,609
Poland	National	1990	0.09	4%	1
Norway	National	1991	64.29	62%	1,725
Sweden	National	1991	139.11	40%	2,821
Denmark	National	1992	28.82	40%	593
Slovenia	National	1996	21.45	24%	92
Estonia	National	2000	2.48	3%	3
Latvia	National	2004	5.58	15%	10
British Columbia	Subnational	2008	27.13	70%	1,107
Liechtenstein	National	2008	100.90	26%	4
Switzerland	National	2008	100.90	33%	1,232
Iceland	National	2010	35.71	29%	57
Ireland	National	2010	24.80	49%	552
Ukraine	National	2011	0.02	71%	4
Japan	National	2012	2.74	68%	2,487
United Kingdom	National	2013	25.46	23%	1,145
France	National	2014	55.30	35%	9,551
Mexico	National	2014	3.01	46%	480
Spain	National	2014	24.80	3%	217
Portugal	National	2015	8.49	29%	171
Alberta	Subnational	2017	23.25	42%	1,080
Chile	National	2017	5.00	39%	145
Colombia	National	2017	5.67	24%	270

Source: World Bank Group (2018).

a. GHG = greenhouse gas emissions. Argentina, Singapore, and South Africa are scheduled to enact carbon taxes in 2019. The carbon tax rate reported is the main rate as of January 2018 reported in dollars. Revenue is an estimate for 2018. The share of emissions covered by the tax is as of January 1, 2019.

Table A2. Canada Province Regressions Data Sources

Variable	*Description*	*Source*
CO_2	Energy-related carbon dioxide emissions	Environment and Climate Change 2018 National Inventory Report (NIR), IPCC-Table C province and territory emissions. Downloaded from http://data.ec.gc.ca/data/substances/monitor/canada-s-official-greenhouse-gas-inventory/C-Tables-IPCC-Sector-Provinces-Territories/?lang=en.
GDP	Gross domestic product	Statistics Canada Table 36-10-0222-01. Expenditure-based GDP in chained $2007. Downloaded from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610022201.
Pop	Population, as of July 1	Statistics Canada Table 17-10-0005-01. Downloaded from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000501.
Export Price	Price index for exports to other countries	Statistics Canada Table 36-10-0223-01. Downloaded from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610022301. Chained $2007.
Employment Shares	Share of full-time workers by industry	Statistics Canada Table 14-10-0023-01. Downloaded from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410002301.
Carbon Tax Rate	Province-level carbon price	BC carbon tax rate from BC Ministry of Small Business and Revenue at https://web.archive.org/web/20130513055926/http:/www.rev.gov.bc.ca/ documents_library/notices/British_Columbia_Carbon_Tax.pdf. AL Specified Gas Emitters Regulation (SGER) price from AL Ministry of Finance documents and set at C$15 per ton of CO2 post-2007. QC carbon price based on average price of QC cap and trade allowance auctions at http://www.environnement.gouv.qc.ca/changements/carbone/ventes-encheres/avis-resultats-en.htm.

Table A3. EU Country Regressions Data Sources

Variable	*Description*	*Source*
GDP	Gross domestic product	OECD data from https://data.oecd.org/gdp/gross-domestic-product-gdp.htm.
Carbon Tax Indicator	Indicator for presence of carbon tax	Data from World Bank (2018).
Emissions Share	Share of GHG emissions covered by carbon tax	World Bank Carbon Pricing Dashboard, https://carbonpricingdashboard.worldbank.org/map_data.

References

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, Ufuk Akcigit, Douglas Hanley, and William Kerr. 2016. "Transition to Clean Technology." *Journal of Political Economy* 124, no. 1: 52–104.

Aldy, Joseph E. 2017. "Designing and Updating a U.S. Carbon Tax in an Uncertain World." *Harvard Environmental Law Review Forum* 41: 28–40.

Aldy, Joseph E. Forthcoming. "Carbon Tax Review and Updating." *Review of Environmental Economics and Policy.* https://www.rff.org/publications/working-papers/designing-and-updating-a-us-carbon-tax-in-an-uncertain-world/.

Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review* 104, no. 10: 3003–37.

Anderson, Soren T., and James M. Sallee. 2016. "Designing Policies to Make Cars Greener." *Annual Review of Resource Economics* 8: 157–80.

Baker, James A., III, Martin Feldstein, Ted Halstead, N. Gregory Mankiw, Henry M. Paulson, Jr., George P. Schultz, Thomas Stephenson, and Rob Walton. 2017. *The Conservative Case for Carbon Dividends*. Washington: Climate Leadership Council.

Barron, Alexander R., Allen A. Fawcett, Marc A. C. Hafstead, James McFarland, and Adele C. Morris. 2018. "Policy Insights from the EMF 32 Study on U.S. Carbon Tax Scenarios." *Climate Change Economics* 9, no. 1: 1–47. doi: https://doi.org/10.1142/S201000781840002X.

Boomhower, Judson, and Lucas W. Davis. 2014. "A Credible Approach for Measuring Inframarginal Participation in Energy Efficiency Programs." *Journal of Public Economics* 113: 67–79.

Borenstein, Severin, and Lucas W. Davis. 2016. "The Distributional Effects of U.S. Clean Energy Tax Credits." In *Tax Policy and the Economy*, edited by Jeffrey R. Brown. University of Chicago Press for National Bureau of Economic Research.

Bovenberg, A. Lans, and Lawrence H. Goulder. 1996. "Optimal Environmental Taxation in the Presence of Other Taxes: General Equilibrium Analysis." *American Economic Review* 86, no. 4: 985–1000.

———. 2002. "Environmental Taxation and Regulation." In *Handbook of Public Economics*, vol. 3, edited by Alan J. Auerbach and Martin Feldstein. Amsterdam: Elsevier Science.

Bovenberg, A. Lans, and Ruud A. de Mooij. 1994. "Environmental Levies and Distortionary Taxation." *American Economic Review* 84, no. 4: 1085–89.

Brännlund, Runar, and Ing-Marie Gren, eds. 1999. *Green Taxes: Economic Theory and Empirical Evidence from Scandinavia*. New Horizons in Environmental Economics. Cheltenham, U.K.: Edward Elgar.

Brännlund, Runar, Tommy Lundgren, and Per-Olov Marklund. 2014. "Carbon Intensity in Production and the Effects of Climate Policy: Evidence from Swedish Industry." *Energy Policy* 67: 844–57.

British Columbia Ministry of Finance. 2019. *Budget and Fiscal Plan 2019/20–2021/22*. https://www.bcbudget.gov.bc.ca/2019/pdf/2019_budget_and_fiscal_plan.pdf.

Bruvoll, Annegrete, and Bodil Merethe Larsen. 2004. "Greenhouse Gas Emissions in Norway: Do Carbon Taxes Work?" *Energy Policy* 32, no. 4: 493–505.

Carl, Jeremy, and David Fedor. 2016. "Tracking Global Carbon Revenues: A Survey of Carbon Taxes versus Cap-and-Trade in the Real World." *Energy Policy* 96: 50–77.

Carlson, Curtis, Dallas Burtraw, Maureen Cropper, and Karen L. Palmer. 2000. "Sulfur Dioxide Control by Electric Utilities: What Are the Gains from Trade?" *Journal of Political Economy* 108, no. 6: 1292–1326.

Carlson, Curtis, and Gilbert E. Metcalf. 2008. "Energy Tax Incentives and the Alternative Minimum Tax." *National Tax Journal* 61, no. 3: 477–91.

Center for Climate and Energy Solutions. No date. "Federal Vehicle Standards." https://www.c2es.org/content/regulating-transportation-sector-carbon-emissions.

Chaffin, Joshua. 2011. "Cyber-Theft Halts EU Emissions Trading." *Financial Times*, January 19. https://www.ft.com/content/27ee8cb0-2401-11e0-bef0-00144feab49a.

Chandra, Ambarish, Sumeet Gulati, and Milind Kandlikar. 2010. "Green Drivers or Free Riders? An Analysis of Tax Rebates for Hybrid Vehicles." *Journal of Environmental Economics and Management* 60, no. 2: 78–93.

Clarke, Leon E., Allen A. Fawcett, John P. Weyant, James McFarland, Vaibhav Chaturvedi, and Yuyu Zhou. 2014. "Technology and U.S. Emissions Reductions Goals: Results of the EMF 24 Modeling Exercise." *Energy Journal* 35, no. S11: 9–31.

Coase, Ronald H. 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3: 1–44.

Cronin, Julie-Anne, Don Fullerton, and Steven E. Sexton. 2017. "Vertical and Horizontal Redistributions from a Carbon Tax and Rebate." Working Paper 23250. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w23250.pdf.

Dales, J. H. 1968. *Pollution, Property, and Prices*. University of Toronto Press.

Eilperin, Juliet. 2017. "EPA's Pruitt Signs Proposed Rule to Unravel Clean Power Plan." *Washington Post*, October 10. https://www.washingtonpost.com/politics/epas-pruitt-signs-proposed-rule-to-unravel-clean-power-plan/2017/10/10/96c83d2c-add2-11e7-a908-a3470754bbb9_story.html?utm_term=.54955d8f72c8.

Elgie, Stewart. 2014. "British Columbia's Carbon Tax Shift: An Environmental and Economic Success." In *Development in a Changing Climate*. Blog, World Bank, Washington. https://blogs.worldbank.org/climatechange/british-columbia-s-carbon-tax-shift-environmental-and-economic-success.

Elgie, Stewart, and Jessica McClay. 2013. *BC's Carbon Tax Shift after Five Years: Results*. Ottawa: Sustainable Prosperity. https://institute.smartprosperity.ca/sites/default/files/publications/files/BC%27s%20Carbon%20Tax%20Shift%20after%205%20Years%20-%20Results.pdf.

Ellerman, A. Denny, Paul L. Joskow, Richard Schmalensee, Juan-Pablo Montero, and Elizabeth M. Bailey. 2000. *Markets for Clean Air: The U.S. Acid Rain Program*. Cambridge University Press.

EIA (U.S. Energy Information Administration). 2018. *Annual Energy Outlook 2018*. Washington: U.S. Government Publishing Office. https://www.eia.gov/outlooks/archive/aeo18/.

EPA (U.S. Environmental Protection Agency). 2018. *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2016*. Washington: U.S. Environmental Protection Agency. https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2016.

Fawcett, Allen, Jim McFarland, Adele Morris, and John Weyant, eds. 2018. "The EMF 32 Study on U.S. Carbon Tax Scenarios." Special issue of the journal *Climate Change Economics*. https://www.worldscientific.com/toc/cce/09/01.

Fischer, Carolyn. 2010. "Renewable Portfolio Standards: When Do They Reduce Price?" *Energy Journal* 31, no. 1: 101–19.

Fullerton, Don, Garth Heutel, and Gilbert E. Metcalf. 2011. "Does the Indexing of Government Transfers Make Carbon Pricing Progressive?" *American Journal of Agricultural Economics* 94, no. 2: 347–53.

Goulder, Lawrence H., Marc A. C. Hafstead, GyuRim Kim, and Xianling Long. Forthcoming. "Impacts of a Carbon Tax across U.S. Household Income Groups: What Are the Equity-Efficiency Trade-Offs?" *Journal of Public Economics*. https://www.nber.org/papers/w25181.

Goulder, Lawrence H., and Andrew R. Schein. 2013. "Carbon Taxes versus Cap and Trade: A Critical Review." *Climate Change Economics* 4, no. 3: 1–28.

Gray, Wayne B., and Gilbert E. Metcalf. 2017. "Carbon Tax Competitiveness Concerns: Assessing a Best Practices Carbon Credit." *National Tax Journal* 70, no. 2: 447–68.

Gruenspecht, Howard K. 1982. "Differentiated Regulation: The Case of Auto Emissions Standards." *American Economic Review* 72, no. 2: 328–31.

Hafstead, Marc, Gilbert E. Metcalf, and Roberton C. Williams III. 2017. "Adding Quantity Certainty to a Carbon Tax through a Tax Adjustment Mechanism for Policy Pre-Commitment." *Harvard Environmental Law Review* 41: 41–57.

Harrison, Kathryn. 2013. "The Political Economy of British Columbia's Carbon Tax." OECD Working Paper ENV/WKP(2013)10. Paris: Organization for Economic Cooperation and Development.

Heal, Geoffrey. 2017. "Reflections: What Would It Take to Reduce U.S. Greenhouse Gas Emissions 80 Percent by 2050?" *Review of Environmental Economics and Policy* 11, no. 2: 319–35.

Hoel, Michael, and Larry Karp. 2002. "Taxes versus Quotas for a Stock Pollutant." *Resource and Energy Economics* 24, no. 4: 367–84.

Horowitz, John, Julie-Anne Cronin, Hannah Hawkins, Laura Konda, and Alex Yuskavage. 2017. *Methodology for Analyzing a Carbon Tax*. Washington: U.S. Department of the Treasury. https://www.treasury.gov/resource-center/tax-policy/tax-analysis/Documents/WP-115.pdf.

Houde, Sébastien, and Joseph E. Aldy. 2017. "Consumers' Response to State Energy Efficient Appliance Rebate Programs." *American Economic Journal: Economic Policy* 9, no. 4: 227–55.

Hsiang, Solomon, Robert Kopp, Amir Jina, James Rising, Michael Delgado, Shashank Mohan, D. J. Rasmussen, Robert Muir-Wood, Paul Wilson, Michael Oppenheimer, Kate Larsen, and Trevor Houser. 2017. "Estimating Economic Damage from Climate Change in the United States." *Science* 356, no. 6345: 1362–69.

International Energy Agency. 2018a. *Energy Prices and Taxes: Second Quarter 2018*. Paris: International Energy Agency.

———. 2018b. *The Future of Cooling*. Paris: International Energy Agency. https://webstore.iea.org/the-future-of-cooling.

Jacobsen, Mark R. 2013. "Evaluating U.S. Fuel Economy Standards in a Model with Producer and Household Heterogeneity." *American Economic Journal: Economic Policy* 5, no. 2: 148–87. doi: 10.1257/pol.5.2.148.

Joskow, Paul L., and Jean Tirole. 2005. "Merchant Transmission Investment." *Journal of Industrial Economics* 53, no. 2: 233–64.

Kaplow, Louis. 1996. "The Optimal Supply of Public Goods and the Distortionary Cost of Taxation." *National Tax Journal* 49 , no. 4: 513–33.

Karp, Larry, and Christian Traeger. 2018. "Prices versus Quantities Reassessed." CESifo Working Paper 7331. Munich: CESfiso Group. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3338660.

Karp, Larry, and Jiangfeng Zhang. 2005. "Regulation of Stock Externalities with Correlated Abatement Costs." *Environmental and Resource Economics* 32, no. 2: 273–99.

Karplus, Valerie J., Sergey Paltsev, Mustafa Babiker, and John M. Reilly. 2013. "Should a Vehicle Fuel Economy Standard Be Combined with an Economy-Wide Greenhouse Gas Emissions Constraint? Implications for Energy and Climate Policy in the United States." *Energy Economics* 36: 322–33. doi: 10.1016/j.eneco.2012.09.001.

Koch, Nicolas, Sabine Fuss, Godefroy Grosjean, and Ottmar Edenhofer. 2014. "Causes of the EU ETS Price Drop: Recession, CDM, Renewable Policies or a Bit of Everything?—New Evidence." *Energy Policy* 73: 676–85.

Leach, Andrew. 2012. "Policy Forum: Alberta's Specified Gas Emitters Regulation." *Canadian Tax Journal* 60, no. 4: 881–98.

Lehane, Bill. 2011. "Hackers Steal Carbon Credits." *Prague Post*, January 26. https://www.praguepost.com/business/7290-hackers-steal-carbon-credits.html.

Lewis, Mark. 2018. "EU Carbon Allowance Market to Shake Its Oversupply Problem." *Financial Times*, April 26. https://www.ft.com/content/15f8e290-46d0-11e8-8ee8-cae73aab7ccb.

Lin, Boqiang, and Xuehui Li. 2011. "The Effect of Carbon Price on per Capita CO_2 Emissions." *Energy Policy* 39, no. 9: 5137–46.

Martin, Ralf, Laurie B. de Preux, and Ulrich J. Wagner. 2014. "The Impact of a Carbon Tax on Manufacturing: Evidence from Microdata." *Journal of Public Economics* 117: 1–14.

McFarland, James R., Allen A. Fawcett, Adele C. Morris, John M. Reilly, and Peter J. Wilcoxen. 2018. "Overview of the EMF 32 Study on U.S. Carbon Tax Scenarios." *Climate Change Economics* 9, no. 1: 1–37.

Metcalf, Gilbert E. 1999. "A Distributional Analysis of Green Tax Reforms." *National Tax Journal* 52, no. 4: 655–81.

———. 2009. "Market-Based Policy Options to Control U.S. Greenhouse Gas Emissions." *Journal of Economic Perspectives* 23, no. 2: 5–27.

———. 2016. "A Conceptual Framework for Measuring the Effectiveness of Green Fiscal Reforms." *International Journal on Green Growth and Development* 2, no. 2: 87–125.

———. 2017. *Implementing a Carbon Tax*. Washington: Resources for the Future. https://www.rff.org/publications/reports/implementing-a-carbon-tax/.

———. 2018. "The Tax Treatment of Oil and Gas Production in the United States: Assessing the Options." *Journal of the Association of Environmental and Resource Economists* 5, no. 1: 1–37.

———. 2019. *Paying for Pollution: Why a Carbon Tax Is Good for America*. New York: Oxford University Press.

———. Forthcoming. "An Emissions Assurance Mechanism: Adding Environmental Certainty to a Carbon Tax." *Review of Environmental Economics and Policy*.

Metcalf, Gilbert E., and David Weisbach. 2009. "The Design of a Carbon Tax." *Harvard Environmental Law Review* 33, no. 2: 499–556.

Murray, Brian, William A. Pizer, and Christina Reichert. 2017. "Increasing Emissions Certainty under a Carbon Tax." *Harvard Environmental Law Review Forum* 41: 14–27.

National Academy of Sciences. 2019. *Negative Emissions Technologies and Reliable Sequestration: A Research Agenda*. Washington: National Acadamies Press.

Newell, Richard G., and William A. Pizer. 2003. "Regulating Stock Externalities under Uncertainty." *Journal of Environmental Economics and Management* 45, no. 2: 416–32.

NHTSA (National Highway Traffic Safety Administration). 2011. "President Obama Announces Historic 54.5 MPG Fuel Efficiency Standard." https://web.archive.org/web/20130305181919/http://www.nhtsa.gov/About+NHTSA/Press+Releases/2011/President+Obama+Announces+Historic+54.5+mpg+Fuel+Efficiency+Standard.

———. 2018. "Corporate Average Fuel Economy." https://www.nhtsa.gov/laws-regulations/corporate-average-fuel-economy.

Nordhaus, William D. 2013. *The Climate Casino: Risk, Uncertainty, and Economics for a Warming World*. New Haven, Conn.: Yale University Press.

———. 2015. "Climate Clubs: Overcoming Free-Riding in International Climate Policy." *American Economic Review* 105, no. 4: 1339–70.

Nordhaus, William D., and Andrew Moffat. 2017. "A Survey of Global Impacts of Climate Change: Replication, Survey Methods, and a Statistical Analysis." Working Paper 23646. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w23646.pdf.

North Carolina Clean Energy Technology Center. 2018. "Renewable Portfolio Standard." http://programs.dsireusa.org/system/program/detail/479.

OMB (U.S. Office of Management and Budget). 2019. *President's Budget Submission to Congress for Fiscal Year 2020.* Washington: OMB. https://www.whitehouse.gov/omb/budget/.

Parry, Ian W. H. 1995. "Pollution Taxes and Revenue Recycling." *Journal of Environmental Economics and Management* 29, no. 3: S64–S77.

Petit, J., R. Jouzel, D. Raynaud, N. I. Barkov, J. M. Barnola, I. Basile, M. Bender, J. Chappellaz, M. Davis, G. Delaygue, M. Delmotte, V. M. Kotyakov, M. Legrand, V. Y. Lipenkov, C. Lorius, L. Pépin, C. Ritz, E. Saltzman, and M. Stievenard. 1999. "Climate and Atmospheric History of the Past 420,000 Years from the Vostok Ice Core, Antarctica." *Nature* 399, no. 6735: 429–36. doi: 10.1038/20859.

Pigou, A. C. 1920. *The Economics of Welfare*. London: Macmillian.

Pindyck, Robert S. 2017. "Coase Lecture: Taxes, Targets and the Social Cost of Carbon." *Economica* 84, no. 335: 345–64. doi: 10.1111/ecca.12243.

Pirttilä, Jukka, and Matti Tuomala. 1997. "Income Tax, Commodity Tax and Environmental Policy." *International Tax and Public Finance* 4, no. 3: 379–93.

Rausch, Sebastian, Gilbert E. Metcalf, and John M. Reilly. 2011. "Distributional Impacts of Carbon Pricing: A General Equilibrium Approach with Micro-Data for Households." *Energy Economics* 33 (supplement 1): S20–S33.

Reagan, Ronald. 1984. "Address before a Joint Session of Congress on the State of the Union." In *The Public Papers of President Ronald W. Reagan*. Washington: U.S. Government Printing Office. http://www.reagan.utexas.edu/archives/speeches/1987/061287d.htm.

Reguant, Mar. 2018. "The Efficiency and Sectoral Distributional Implications of Large-Scale Renewable Policies." Working Paper 24398. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24398.pdf.

Rivers, Nicholas, and Brandon Schaufele. 2015. "Salience of Carbon Taxes in the Gasoline Market." *Journal of Environmental Economics and Management* 74: 23–36.

Sallee, James M. 2011. "The Taxation of Fuel Economy." *Tax Policy and the Economy* 25, no. 1: 1–37.

Schmalensee, Richard, and Robert N. Stavins. 2013. "The SO_2 Allowance Trading System: The Ironic History of a Grand Policy Experiment." *Journal of Economic Perspectives* 27, no. 1: 103–22.

Smale, Robin, Murray Hartley, Cameron Hepburn, John Ward, and Michael Grubb. 2006. "The Impact of CO_2 Emissions Trading on Firm Profits and Market Prices." *Climate Policy* 6, no. 1: 31–48.

Szabo, Mike. 2019. "South African Carbon Tax Bill Heads for Parliament Vote After Committee Approval." Carbon Pulse. https://carbon-pulse.com/68529/.

U.S. Interagency Working Group on the Social Cost of Carbon. 2016. *Technical Support Document: Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis under Executive Order 12866*. Washington: U.S. Environmental Protection Agency. https://www.epa.gov/sites/production/files/2016-12/documents/sc_co2_tsd_august_2016.pdf.

Wagner, Gernot, and Martin L. Weitzman. 2015. *Climate Shock: The Economic Consequences of a Hotter Planet*. Princeton University Press.

Wald, Matthew L. 2012. "An Argument Over Wind." *New York Times*, September 14. https://green.blogs.nytimes.com/2012/09/14/an-argument-over-wind/

Weitzman, Martin L. 1974. "Prices vs. Quantities." *Review of Economic Studies* 41, no. 4: 477–91.

Williams, James H., Benjamin Haley, Fredrich Kahrl, Jack Moore, Andrew D. Jones, Margaret S. Torn, and Haewon McJeon. 2014. *Pathways to Deep Decarbonization in the United States*. U.S. report of Deep Decarbonization Pathways Project of Sustainable Development Solutions Network and Institute for Sustainable Development and International Relations. Revision with Technical Supplment, November 16, 2015. San Francisco: Energy and Environmental Economics. http://unsdsn.org/wp-content/uploads/2014/09/US-Deep-Decarbonization-Report.pdf.

Wingrove, Josh. 2019. "Carbon Tax Takes Effect in Industrial Heartland." *Climate Wire*, January 2. https://www.eenews.net/climatewire/2019/01/02/stories/1060110691.

World Bank Group. 2018. *States and Trends of Carbon Pricing 2018*. Washington: World Bank Group. https://openknowledge.worldbank.org/handle/10986/29687.

World Economic Forum. 2018. *Global Competitiveness Report 2017–2018*. Geneva: World Economic Forum. http://www3.weforum.org/docs/GCR2017-2018/05FullReport/TheGlobalCompetitivenessReport2017%E2%80%932018.pdf.

Yamazaki, Akio. 2017. "Jobs and Climate Policy: Evidence from British Columbia's Revenue-Neutral Carbon Tax." *Journal of Environmental Economics and Management* 83: 197–216.

Comments and Discussion

COMMENT BY

LAWRENCE GOULDER Gilbert Metcalf has produced an outstanding paper for this volume. The paper is impressive along many dimensions. One is scope. The paper offers:

—scientific background on the climate change problem (including information both on historical changes in climate and on scientists' discovery and understanding of the problem);

—the economic rationale for a carbon tax: the theory of how (Pigouvian) taxes such as carbon taxes can produce an efficiency-improving policy response to externalities;

—a range of policy alternatives to a carbon tax, with theoretically and empirically based assessments of the strengths and weaknesses of these options;

—a summary of what economic models have indicated regarding the costs of achieving reductions under alternative policy efforts;

—a review of accomplishments and difficulties associated with climate policy efforts in the United States and other countries; and

—estimates from several econometric studies (including some by Metcalf) of the carbon tax's impact on emissions, GDP, and employment.

Notwithstanding its considerable breadth, the paper's treatment of these topics is not superficial. The analysis gets to the heart of the critical issues, invoking relevant theory and empirical findings, and supporting key points with compelling real-world examples.

Together, these features make the paper as good an introduction to the economics and policy issues surrounding the carbon tax as I have seen anywhere. (I have already assigned the paper to students in one of my courses.)

Here, I focus on three issues connected with Metcalf's paper. First, I consider the question: As an instrument for emissions pricing, how attractive is a carbon tax relative to its chief competitor—cap and trade? Second, motivated by recent scientific evidence that strong action to reduce emissions of carbon dioxide (CO_2) is urgent, I consider the implications of urgency for the choice among climate policy alternatives. Finally, I consider the extent to which a carbon tax needs to be accompanied by direct promotion of "breakthrough" low-carbon technologies, and what that might mean for how policymakers might employ a carbon tax.

THE CARBON TAX VERSUS THE COMPETITION Metcalf's paper offers sound arguments as to why implementing a carbon tax might achieve target reductions in emissions of CO_2 at a lower cost than direct regulation (for example, mandated technologies) and subsidies. It also argues that the carbon tax is a better choice than cap and trade, the principal emissions-pricing alternative to a carbon tax. The paper argues that the carbon tax has three key advantages over cap and trade: (1) it entails less administrative complexity, (2) it escapes important problematic interactions with other environmental policies, and (3) it avoids (to a significant degree) emissions-price uncertainty and fluctuations in emissions prices.

For each of these three considerations, the paper provides compelling detail. Regarding the first, the paper indicates that cap and trade requires the regulatory authority not only to keep track of covered facilities' emissions (a requirement under the carbon tax as well) but also a "new administrative structure to create allowances, track the hold auctions or otherwise distribute them and develop rules to avoid fraud or abuse." Regarding the second, the paper describes how policy interactions have caused difficulties in the European Union's Emissions Trading System (ETS) and in the East Coast's Regional Greenhouse Gas Initiative. Under cap and trade, other regulations can interfere with cap and trade by affecting demand and supply for allowances and the equilibrium allowance prices. This is not a problem for the carbon tax, because tax rates are set (fixed) by the government. Regarding the third, the paper refers to difficulties associated with varying allowance prices in the ETS and in the U.S. Environmental Protection Agency's Acid Rain Program.

These are important arguments. In the United States in recent years, a carbon tax seems to have gained popularity relative to cap and trade as an option for federal-level climate policy. The "third advantage" mentioned above—that a carbon tax avoids uncertainties and fluctuations in emissions prices—seems to explain much of this development. The observed fluctuations of allowance prices in the ETS and the Environmental Protection Agency's Acid Rain Program have soured many analysts on cap and trade.

That said, it should be recognized that a carbon tax comes with its own form of uncertainty. It implies uncertainty about emissions *quantities*. The emissions levels that will result under a given carbon tax program are not specified in advance but rather are determined by producers' responses to the tax. This contrasts with cap and trade, which leaves little uncertainty about emissions quantities, assuming good enforcement of the emissions limits implied by the number of emissions allowances in circulation. An especially important consideration in deciding between these two options for emissions pricing is the relative cost of cap and trade's emissions price uncertainty and the carbon tax's emissions quantity uncertainty. I would have given more attention to the emissions uncertainty issue than is offered in the paper. Still, I tend to find persuasive Metcalf's overall conclusion about the relative attractiveness of a carbon tax. But both options have significant advantages over conventional regulation. Adoption of either of these price-based instruments at the national level would be a major step forward for U.S. climate change policy.

THE URGENCY OF STRONGER CLIMATE POLICY ACTION JUSTIFIES ATTACHING GREATER WEIGHT TO POLITICAL FEASIBILITY Metcalf's paper identifies several important criteria relevant to the evaluation of a carbon tax, including cost-effectiveness, fairness, administrative ease, and political feasibility. Because the assignment of weights is inherently subjective, it is understandable that the paper avoids recommending how much weight to attach to each of these criteria. Nevertheless, I think it is vitally important to recognize that political feasibility is becoming especially important in view of the scientific findings that a delay in taking strong action on climate change will be very costly. As I indicate here, giving greater weight to this dimension can affect policy rankings.

Over the past decade, the consensus scientific findings about the potential extent of future climate changes and their biophysical consequences have become increasingly ominous. Climate scientists often focus on the potential biophysical outcomes associated with given increases (relative to preindustrial levels) in global average surface temperature. One focal point has been an increase of 2 degrees Celsius. Twelve years ago, a synthesis report from the Intergovernmental Panel on Climate Change (IPCC 2007) indicated that a 2-degree increase would lead to substantial climate change and very serious associated biophysical effects. The most recent comparable report (IPCC 2018) indicates that the effects of a 2-degree increase would be considerably more severe. A 1.5-degree increase is now considered sufficient to produce climate-related damage of comparable magnitude to those previously attributed to a 2-degree increase. The most recent

IPCC report indicates that, with 50 percent probability, the atmospheric concentrations that would produce a 1.5-degree temperature increase would be reached in 10 to 20 years if the current global rate of emissions of CO_2 were to continue. To me, this implies urgency. Of course, international efforts can reduce the global rate. But my own calculations suggest that full compliance with the commitments under the 2015 Paris Accord would extend the time window only by about 10 percent; that is, the 1.5-degree temperature increase would be reached in 11 to 22 years (Goulder 2019).

Under these circumstances, a delay in achieving significant reductions in emissions of CO_2 is costly. Relative to a scenario involving nearer-term action, a delay implies faster increases in atmospheric concentrations of CO_2, more extensive climate change (including increased average global surface temperature), and more serious damage related to climate change. An alternative way to view the cost of a delay is to consider the cost of preventing atmospheric concentrations from exceeding some particular concentration level that is deemed unacceptable. In this context, a delay necessitates accelerated future reductions in emissions to prevent atmospheric concentrations from exceeding that level. Assuming rising marginal costs of abatement, the accelerated reductions might be extremely costly.[1]

Political feasibility is always worthy of consideration; but in the climate change context, it takes on greater importance because it connects with the cost of a delay. A policy with greater political prospects—that is, a greater chance of near-term implementation—implies lower expected climate-related damage than a policy with more meager political prospects, other things being equal. Suppose that, conditional on implementation at a given point in time, policy A achieves some given emissions reduction target at a lower cost than policy B. But suppose that policy B has a much greater chance of implementation in the near term. Then the *expected* cost of policy B could be lower than that of policy A.[2] Policy B's earlier implementation would avoid some of the cost of a delay.

1. One offsetting benefit from delay is that it allows time for discovery of new and lower-cost methods for emissions abatement. On this, see, for example, Jaffe, Newell, and Stavins (2003). It seems impossible to quantify the extent to which this benefit offsets the additional risks posed by delay. Still, the potential for severe climate-related costs from delay seems to justify the assumption that delay is quite costly overall.

2. An alternative accounting method yields the same result. Instead of referring to the higher environmental damage as a greater cost of policy A, we can view both policies as having the same (more narrowly defined) cost, while indicating that policy B yields larger environmental benefits (avoided climate damage). In this case, policy B is again preferred because its net benefits are higher.

Such considerations of political feasibility can affect the relative expected costs of alternative climate policy options. Consider the class of revenue-neutral carbon tax policies. Within this category, the options include (1) recycling in the form of lump-sum cash rebates and (2) recycling via cuts in the marginal rates of corporate or individual income taxes. Economists typically view the policies with marginal rate cuts as more cost-effective, because reducing marginal rates reduces the excess burden of such taxes. Numerical simulations support these perspectives.[3] However, if political support (and the odds of near-term implementation) is much higher for lump-sum recycling, the expected policy cost could in fact be lower. This is not meant to declare that the expected cost is clearly lower under lump-sum recycling. But it is meant to urge consideration, in comparing policy options, of the cost implications associated with political feasibility. This can affect the cost rankings in important ways.

These considerations also motivate revisiting our assumptions about the relative cost-effectiveness of alternatives to a carbon tax. Some analysts claim that a nationwide clean energy standard has better political prospects than a carbon tax, in part because its costs seem to be less salient than the costs associated with a carbon tax. Consequently, although studies suggest that it may have a disadvantage according to a narrower cost-effectiveness measure—one that does not account for prospects for near-term implementation—it could potentially emerge as less costly once such prospects are considered.[4] Given the very high stakes of the climate change problem in relation to future human welfare, as well as the urgency of action, the potential political prospects of this policy deserve consideration as part of the overall cost assessment. This policy might deserve a better rating than it is often given. Likewise, it seems worth employing this framework to reinvestigate the overall costs of achieving reductions

3. For example, Marc Hafstead and I have applied our intertemporal general equilibrium model to assess the effects of a broad-based U.S. carbon tax implemented in 2017, reaching \$20 per ton after a three-year phase-in, and increasing at 4 percent a year in real terms. As reported by Goulder and Hafstead (2017), we find that over the period 2017–50, the welfare costs per ton of reduced CO_2 are about \$42 when revenues are recycled through lump-sum rebates, as compared with about \$31 when recycling is via cuts in individual income taxes. The numbers for welfare costs are according to the equivalent variation measure over the 2017–50 time interval.

4. My paper with Marc Hafstead and Roberton Williams (2016) finds that, ignoring probabilities of implementation, a clean energy standard that achieves moderate or large reductions in emissions is less cost-effective than an equally stringent carbon tax. However, it is slightly more cost-effective at low stringency levels. This stems from the clean energy standard's ability to avoid the certain price increases that distort factor markets.

via subsidies to CO_2 abatement. I am not claiming that these alternatives are better than the carbon tax, but I believe it is worth considering them, along with a carbon tax, with attention to their political prospects.

Readers might feel that political feasibility is beyond the purview of economists, and that, accordingly, economists should not aim to incorporate relative likelihood of near-term implementation in their assessments of policy alternatives. I do not mean to suggest that economists become political scientists. However, economists can nevertheless incorporate considerations of timing into their analyses. To assess the potential savings that policy A might have over policy B as a result of better prospects for near-term implementation, one would need, for each policy, (1) subjective probabilities of implementation at various points in time in the future, along with (2) estimates of the differences in expected climate damage from the two policies, with the estimates being a function of the differences in implementation probabilities at various points in time. It is well within the domain of economic analysis to translate this information into expected cost savings. The subjective probabilities could be elicited from politicians and political scientists; the differences in expected damage would be elicited from climate scientists. Obviously, different experts would offer different numbers. Nonetheless, the resulting framework would provide valuable information by making explicit what needs to be assumed about implementation probabilities and avoided climate damage to make one given policy's overall costs lower than another's. This would help focus the debates about the relative attractiveness of the policies under consideration.

THE INTERCONNECTED ROLES OF TECHNOLOGY POLICY AND THE CARBON TAX In evaluating the various policy options, Metcalf's paper focuses mostly on "emissions-oriented" policies—policies that aim to reduce emissions by providing incentives or requirements for fuel-switching, end-of-pipe treatment, or conservation (reduced product demand). The paper makes clear, however, that there is also a role for policies that directly promote the discovery and development of new technologies. In this connection, it points out that, in addition to the market failure from the externality associated with emissions, there is an *innovation market failure* stemming from a beneficial externality, the knowledge that is not appropriated by the inventor and spills over to other producers. This additional market failure yields a rationale for combining an emissions-oriented policy (such as a carbon tax) with public policy to augment producers' incentives for research efforts. By introducing policies to increase incentives for research, the government addresses the beneficial spillover externality and thereby helps raise levels of research effort to a socially more efficient level.

In discussing the need for new technologies, Metcalf's paper mentions that "international climate negotiators have focused on a global goal of reducing emissions by 80 percent relative to 2005 by 2050. . . . Most economic analyses suggest that given the current state of technological progress, an 80 percent reduction by 2050 would be extremely costly." He refers to results from a 2014 model comparison study by Stanford's Energy Modeling Forum, which indicate that this percentage reduction would be somewhere in the range of $200 to over $500 a ton.[5] As Metcalf notes, the carbon price needed to reach any future emissions target will depend to a large degree on the pace of clean energy technological development.

We cannot tell in advance what low-carbon technologies will emerge, and what will be the costs per ton of the emissions reductions they bring about. What should we do in the face of this uncertainty? The paper concentrates on approaches in which the government first sets a carbon tax time profile and subsequently adjusts the profile of tax rates as needed to keep the United States on a path to reach its long-run emissions reduction goals. One could refer to this as the "fixed-target" approach. An alternative approach is to set a time profile of carbon tax rates, based on estimates of the social cost of carbon, and let the emissions reduction outcome be determined by this carbon tax time profile based on this social cost. This latter approach is what considerations of economic efficiency would recommend.[6]

The fixed-target approach might have the edge in terms of political acceptability. Environmental groups, in particular, often prefer to establish emissions reduction targets over the simple establishment of an emissions price (carbon tax) profile, because the latter would not assure particular outcomes in terms of emissions reductions. However, it is worth noting that any particular quantity target could end up implying marginal costs per ton of emissions abatement that are very different from the social cost of carbon (which is problematic in terms of efficiency) and require exceptionally high carbon tax rates to achieve convergence to the target. The

5. This is Energy Modeling Forum Study 24. The results of this study are described by Clarke and others (2014).

6. This assumes that within the relevant range, the marginal damage schedule is relatively flat compared with the marginal abatement cost schedule. As Weitzman (1974) has shown, under such circumstances the expected net benefits are greater under a price-based policy than a policy in which the aggregate quantity is fixed.

fixed-target approach does have an adjustment mechanism—tax rates would be changed as new information arises—but note that these adjustments do not assure that astronomical tax rates are avoided. This approach gives little focus to whether the carbon prices needed to achieve convergence are much too high or much too low from an efficiency point of view. Rather, the adjustments to tax rates are whatever are needed to help the cumulative emissions reductions converge on the ultimate target. I doubt that, in the future, politicians would be willing to stand behind an adjustment mechanism that would require extremely high carbon tax rates to bring about convergence.

This suggests the value of an alternative approach—one with an adjustment mechanism that balances the goals of (1) coming close to initial emissions reduction targets and (2) employing carbon tax rates not far from the estimated social cost of carbon. I would have liked to see some discussion of this alternative. Of course, this alternative approach has the handicap of being more complicated. Also, it might have less political appeal, at least initially.[7]

Apart from political considerations, the most compelling approach, in my view, is a policy process whereby the government simply sets the time profile of carbon tax rates, based on the (central) estimates of the social cost of carbon, and adjusts the profile over time as new information on the social cost of carbon arrives. No quantity targets would be employed under this approach, which would be the most efficient one.

FINAL COMMENTS There is a strong consensus among climate scientists that in the absence of significant reductions in greenhouse gas emissions from anticipated business-as-usual levels, future climate change will be extensive and cause substantial harm to humans and other species. Metcalf's paper provides an outstanding overview of the economics of climate change policy. It identifies a range of policy options that the United States could employ for addressing this important problem, and it adroitly combines theory and empirical evidence to reveal the strengths and weaknesses of each option. It argues that a carbon tax has important advantages over the other options.

The paper is a great source for anyone wishing to become better acquainted with the economics of climate change policy and the relative

7. "Initially" is important here because the fixed-target approach would likely lose popularity if converging to the target eventually required extremely high carbon tax rates. This could happen if the emissions reductions from newly discovered low-carbon technologies turned out to be very meager.

advantages and disadvantages of important policy options in terms of cost, distributional effects, and political acceptability.

My comments do not contradict Metcalf's key conclusions, but rather bring out issues that I believe deserve further close attention. I have supplemented Metcalf's key points with (1) a further focus on a particular attraction of cap and trade (less uncertainty about policy-induced emissions reductions), (2) the urgency of more stringent climate change policy and its implications for policy rankings, and (3) the choice between setting a carbon tax based on given targets for cumulative emissions reductions versus setting the carbon tax time profile and letting the cumulative reductions be endogenous. I am especially concerned about item 2. Urgency justifies giving considerable weight to the probabilities of near-term implementation in the evaluation of climate policy alternatives. Doing so can lead to different rankings of policy costs and thereby affect the types of policies that economists endorse. As I have indicated here, I believe that economists can incorporate considerations of political feasibility within a strictly economic evaluation framework—that is, without straying from their domain of expertise.

REFERENCES FOR THE GOULDER COMMENT

Clarke, Leon E., Allen A. Fawcett, John P. Weyant, James McFarland, Vaibhav Chaturvedi, and Yuyu Zhou. 2014. "Technology and U.S. Emissions Reductions Goals: Results of the EMF 24 Modeling Exercise." *Energy Journal* 35, no. S11: 9–31.

Goulder, Lawrence H. 2019. "Addressing the Urgency of More Stringent Climate-Change Policy." Working Paper 19-12. Washington: Resources for the Future.

———. Forthcoming. "Addressing the Urgency of More Stringent Climate-Change Policy." *Review of Environmental Economics and Policy* 14, no. 1.

Goulder, Lawrence H., and Marc A. C. Hafstead. 2017. *Confronting the Climate Challenge: Options for U.S. Policy.* New York: Columbia University Press.

Goulder, Lawrence H., Marc A. C. Hafstead, and Roberton C. Williams III. 2016. "General Equilibrium Impacts of a Federal Clean Energy Standard." *American Economic Journal: Economic Policy* 8, no. 2: 186–218.

IPCC (Intergovernmental Panel on Climate Change). 2007. *Climate Change 2007: Synthesis Report—Contribution of Working Groups I, II, and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by R. K. Pachauri and A. Reisinger. Geneva: IPCC. https://www.ipcc.ch/site/assets/uploads/2018/02/ar4_syr_full_report.pdf.

———. 2018. *Summary for Policymakers of Global Warming of 1.5°C.* Geneva: IPCC. https://www.ipcc.ch/site/assets/uploads/sites/2/2019/05/SR15_SPM_version_report_LR.pdf.

Jaffe, Adam, Richard Newell, and Robert Stavins. 2003. "Technological Change and the Environment." In *Handbook of Environmental Economics, Volume 1*, edited by Karl-Goran Mäler and Jeffrey R. Vincent. Amsterdam: Elsevier Science.

Weitzman, Martin L. 1974. "Prices vs. Quantities." *Review of Economic Studies* 41, no. 4: 477–91.

COMMENT BY

ADELE C. MORRIS This paper by Gilbert Metcalf elucidates the merits of a tax on carbon dioxide (CO_2) and other greenhouse gas (GHG) emissions—a carbon tax, for short. Framing climatic damage from human-induced GHG emissions as a textbook example of external costs, the paper argues that a price on carbon is necessary to any cost-effective climate policy portfolio. It also calls on policymakers to supplement a carbon tax with support for innovation in low-cost low-carbon technologies. The paper ably distills the reasoning and research behind the overwhelming support by economists for a U.S. carbon tax (Climate Leadership Council 2019).

After a brief review of the science of climatic disruption, the paper examines three policy scenarios: regulating GHGs using existing statutory authority; nontax options for new legislation; and alternative implementations of a carbon tax. It also surveys the evidence on the performance of existing carbon tax policies in other countries and presents new analysis of the outcomes of the carbon tax in British Columbia.

A fulsome discussion of climate science is even more beyond the scope of this comment than it is beyond the scope of Metcalf's paper, but two recent syntheses highlight the potentially severe outcomes globally (IPCC 2018) and within the United States (USGCRP 2017). Stipulating that unchecked climate change and ocean acidification are not in the interests of humanity, the focus goes to what to do about it.

A carbon tax works by shifting the relative price of energy sources by an amount that reflects their CO_2 emissions. For example, natural gas has about half the carbon per unit of energy as coal, so its after-tax price will rise less than coal's, inducing substitutions across fossil fuels. The carbon tax does not directly affect the cost of renewable power, making it relatively more economical. These shifts in relative prices immediately drive dispatch in the power sector toward lower-carbon generators. In the longer run, investors have incentives to develop and deploy lower-carbon technologies.

A carbon tax can be straightforward to administer, particularly with a judicious choice of the point in the supply chain where the tax is imposed. The Congressional Research Service (CRS 2019) estimates that a carbon tax could cover about 77 percent of U.S. GHG emissions with fewer than 2,000 taxpayers. For some, the tax could piggyback on existing federal excises, adding little administrative burden. Other gases and sources can be included as feasible.

Some emissions are poorly suited to a carbon tax approach. They may be hard to measure (for example, methane and nitrous oxide from rice cultivation or changes in carbon stored in agricultural soils) or it may be hard to identify a responsible party (certain fugitive emissions from the natural gas system). Some sources, such as aviation fuels, currently lack lower-GHG substitutes, so though a carbon tax incentivizes the long-run development of new technologies, in the short run it produces few emissions benefits. Controlling these emissions over the long run may best be accomplished through nontax policies; this is an important area for more research.

SETTING THE PRICE ON CARBON Metcalf emphasizes the Pigouvian nature of a carbon tax, particularly in suggesting that the tax can be set at an estimate of the climate damage from an incremental ton of CO_2 emissions, a.k.a. the social cost of carbon. Though in principle this is right, in practice numerous complications arise. One is the intractable task of estimating the monetary damage associated with nonmarket outcomes, such as species extinction, disrupted ecosystems, and expanded vector-borne diseases. Challenges also include the choice of the baseline against which to estimate damage, uncertainty in human and natural systems, an incomplete understanding of damage channels, and the discounting and aggregation of effects over time and across widely differing societies (Rothman, Amelung, and Polomé 2003). Further controversies arise over how to account for potentially nonmarginal or threshold damage, such as the disintegration of the West Antarctic Ice Sheet (Diaz and Keller 2016).

One can take as a benchmark the social cost of carbon used by White House agencies for monetizing the GHG effects of regulations. However, even this is a moving target. The Trump administration dramatically lowered the values adopted by the Obama administration, in part by excluding damage outside the United States and raising the rate at which future damage is discounted to current dollars.[1]

1. See U.S. Interagency Working Group on the Social Cost of Greenhouse Gases (2016); and EPA (2018, table 4-1).

Another approach is to set and revise a carbon tax trajectory to hit a particular emissions goal. Economists have offered a number of proposals to do this, involving various degrees of discretionary and formulaic adjustments.[2] Emissions certainty has a particular appeal for environmental advocates, and a focus on emissions goals is consistent with typical pledges under the United Nations Framework on Climate Change (Brooks 2016). On the other hand, including in law measures to adjust carbon tax rates over time, particularly with a formula or third-party determination, strikes me as a heavy legislative lift. Congress rarely even adjusts excise tax rates for inflation. Moreover, the environmental advantage of exactly hitting a specific annual emissions level in one country in a particular year relative to, say, being 5 percent off, is small. Climatic damage derives from the stock of GHGs in the atmosphere, the cumulative result of many decades of global emissions. If unforeseen carbon tax adjustments add uncertainty to the price signal, then it is reasonable to ask whether the costs of emissions certainty mechanisms are justified by their benefits. Certainly, Congress should revisit the policy regularly in light of new information. The question is whether that will be left to future legislators or incorporated into current law.

Regardless of the optimality of any one tax trajectory, the ambition of climate policy in the United States remains importantly bounded by the inclinations of the American electorate. An unduly high carbon price will invite disorderly collapse at the next recession, a change of political party in control of the government, or a spike in oil prices. If investors discount the duration of the policy, the effective price signal will fall below the statutory price, undermining the intended performance of the tax.

The sweet spot between a consensus and overambition is anyone's guess. However, we can take as one example of what not to do from what happened in Australia. In 2012, the government adopted a poorly designed, highly partisan carbon-pricing policy, and the next year a new government promptly ended it (Crowley 2017). Australian GHG emissions have trended upward since 2015, and the issue remains politically contentious.[3] In my view, the downside risks of policy reversals are so costly as to warrant choosing a tax trajectory that endures, ideally with bipartisan support, even if it falls short of a proper Pigouvian price or fails to ensure a particular long-run emissions outcome.

2. For example, see the paper by Murray, Pizer, and Reichert (2017), and also see the other papers in the same issue of the *Harvard Environmental Law Review Forum*.

3. Australian Department of the Environment and Energy (2018, figure 3).

REGULATION AND SUBSIDIES Metcalf deftly describes the drawbacks of climate-related regulatory efforts so far, including tightening automotive fuel economy standards and the Clean Power Plan. Despite the best efforts of Obama's climate team, the Trump administration is dismantling nearly everything they did. Even if lawsuits delay Trump's actions, recent developments serve as a reality check on the potential of existing federal regulatory authorities and executive actions, or those at the state level for that matter, to reduce U.S. emissions over the long run. The climate challenge needs congressional action (Morris and Gross 2018).

Subsidies face some of the same fickle politics as regulation, but at least they create vested interests in their perpetuation. Metcalf cites as an example renewable portfolio standards (RPSs), which reward renewable power generators at the expense of their carbon-intensive competitors. In principle, an RPS could in part mimic the outcomes of a carbon tax, which modeling shows would reduce emissions dramatically and efficiently from the power sector (McFarland and others 2018). Conversely, renewable power is intermittent and requires some sort of backup power, which adds costs (Greenstone and Nath 2019). Second, renewable power plants can incur relatively high costs for land and transmission to far-removed consumers. Third, requiring new renewable capacity can displace existing zero-carbon nuclear power rather than a fossil alternative. Finally, a power-sector-only policy begs the question of how to abate GHGs from industry and transportation. A sector-by-sector approach distorts investment across sectors and sources, ultimately raising the cost of a given level of abatement.

A similar approach to an RPS, a clean energy standard, gives credit to a broader range of lower-carbon generation, such as nuclear and natural gas. If Congress is intent on a power-sector-only policy, a clean energy standard or power-sector-only carbon tax would be superior to an RPS. In addition to promoting renewables, a broader approach prompts fuel-switching from coal to natural gas and helps preserve the economic life of existing nuclear power, which is important for long-run decarbonization.

If policymakers' focus on the power sector derives from a concern about imposing a large-jump discontinuity in gasoline prices, they can adopt an economy-wide approach that taxes all carbon but eases in the price signal on transportation fuels. This achieves the desired short-run low-cost abatement from the power sector while preserving long-run incentives to abate emissions from all sources. As an example, the bill sponsored by Representative Carlos Curbelo (R-Fla.) in 2018 would have eliminated federal taxes on gasoline, diesel, and aviation fuels and replaced them with an economy-wide carbon tax that would increase over inflation each year (Hafstead 2018).[4]

Metcalf reviews the downsides to other GHG-related subsidies, such as production and investment tax credits for renewable power and tax credits for purchases of electric vehicles. These policies are in no way a substitute for an economy-wide carbon price. Less settled is whether some subsidies make sense as interim measures or as complements to a price signal to address the externality in innovation. This is a ripe area for research.

DISADVANTAGES OF CAP AND TRADE RELATIVE TO A CARBON TAX Metcalf's paper reviews the disadvantages of a cap-and-trade system relative to a carbon tax: price volatility, administrative complexity, market uncertainty for innovators, and limiting the environmental benefits of supplementary policies. The last of these is especially important in the context of how federal policy affects the environmental benefits of subfederal policies. Under a federal cap-and-trade program, state and local governments that take on more ambitious climate efforts merely free up allowances for use in other jurisdictions. Under a federal carbon tax, state and local governments can amplify the environmental benefits of the federal excise with whatever additional policies they see fit. This consideration is more important now than it was 10 years ago, when Congress considered a cap-and-trade approach.[5] Since then, state-level climate and energy policies have proliferated, including measures to cap GHG emissions, promote renewables, and invest in energy efficiency.[6] To me, it makes little sense to obviate new gains from these programs by adopting a federal, cap-based approach.

One option to cushion the burden of overlapping policies is for the federal carbon tax policy to give temporary and declining credits to entities that must pay for their GHG emissions at the state level, as reflected in the Curbelo Bill mentioned above. This is a little more complicated than it sounds, however, because the regulated entities at the state level are likely to be downstream from federal carbon taxpayers—that is, not the same firms.

4. Market Choice Act, H.R. 6463, 115th Congress (https://www.congress.gov/bill/115th-congress/house-bill/6463).

5. In 2009, the U.S. House of Representatives passed legislation sponsored by representatives Henry Waxman (D-Calif.) and Edward Markey (D-Mass.) that would have established an economy-wide GHG cap-and-trade system along with other supplementary measures. The effort died in the Senate.

6. See the compendium "U.S. State Climate Action Plans" (Center for Climate and Energy Solutions 2019).

A U.S. CARBON TAX COULD BE A POWERFUL TOOL IN THE GLOBAL CLIMATE CHANGE CHALLENGE Metcalf notes that climate policy in the United States alone cannot contain global concentrations of GHGs and thwart further warming. Therefore, one critical lens through which to assess U.S. policy is the degree to which it would foster abatement abroad. At least three channels of influence could apply, and a carbon tax dominates both cap-and-trade and regulation along each channel. First, to the extent that U.S. policy promotes the development of low-cost technologies, abatement in other countries could be less costly and, by extension, greater. By harnessing the profit motive in the world's largest market, a carbon tax would unleash the ingenuity of American scientists and engineers, supported by the unsurpassed breadth, depth, and liquidity of U.S. capital markets. The technologies forged in U.S. markets, enabled by support for basic research and development from the federal government, could be the greatest contributions the United States makes to the global climate effort.

Second, in contrast to emissions caps and regulations under section 111(d) of the Clean Air Act, the economic effort of a carbon tax, at least on the margin, is clear to all. If the United States adopts a transparent and predictable carbon price, its negotiators can more effectively press other countries for economically comparable commitments. In my view, serious climate policy is serious economic policy, and progress will be slow as long as climate remains in the exclusive domain of relatively weak environment ministries. Reframing climate negotiations as economic negotiations that emphasize mutually agreeable carbon price levels or floors and are led by the more powerful finance ministries could offer a new dynamic for progress (McKibbin, Morris, and Wilcoxen 2014).[7]

Finally, a carbon tax approach can allow the United States to impose import duties on high-GHG goods (Morris 2018). This could motivate other countries to lower the carbon intensity of their exports or negotiate exemptions by demonstrating that they have adopted comparable measures. Border carbon adjustments would be difficult under any climate program, but determining whether other countries' policies are comparable could be more complicated in a cap-and-trade program with volatile allowance prices. Border adjustments would be impossible under current regulatory authority.

7. Some finance ministries have begun to convene on climate action, including "climate informed fiscal policymaking," under the auspices of the World Bank's Climate Action Peer Exchange (CAPE); see World Bank (2019). In full disclosure, I have served on CAPE's technical advisory group.

EMPIRICAL EVIDENCE Metcalf reviews the limited evidence of the outcomes of existing carbon tax policies. In addition to the paucity of the research, we cannot use it reliably to project the likely outcomes within the United States. Each country (or subfederal jurisdiction) that has adopted a carbon tax has had its own idiosyncratic policy design, baseline fuel mix, industrial composition, and low-carbon resource base. What we can say from the few studies available so far is that carbon tax policies appear to have reduced emissions without appreciable economic impedance.

Economic modeling, albeit flawed, is the best tool available to inform U.S. climate policy development. Recent multimodel analyses of a U.S. carbon tax offer several key lessons.[8] First, even a modest carbon tax starting at \$25 per ton of CO_2 and rising gradually over inflation can dramatically reduce U.S. GHG emissions, particularly in the power sector. This outcome primarily derives from a rapid shift away from coal. Coal is the most carbon-intensive fuel, and competes with many lower-carbon substitutes in its primary market of power generation. The robust finding that a carbon tax would dramatically reduce coal production in the United States warrants measures to assist coal workers and coal-reliant communities. The details of the best ways to do this remain for future research.

Along with reducing CO_2, a carbon tax sharply reduces other air pollutants such as sulfur dioxide, nitrogen oxides, mercury, and particulate matter. These reductions would provide significant near-term domestic benefits for human health and the environment.

Another key lesson from the modeling is that the environmental benefits of a carbon tax are not diminished by returning the revenue back to households through rebates or cuts in other taxes. This means that policymakers have great discretion to achieve distributional or other goals without compromising the policy's primary function.

Further, policymakers need not worry about significant effects on GDP growth. Modeling suggests that an efficiently designed, economy-wide tax on carbon produces only minor perturbations in economic growth rates, and that does not account for the economic benefits of a safer climate and cleaner air. For example, models project that a policy that starts at \$25 per ton of CO_2 and rises at 5 percent over inflation results in an average GDP growth rate through 2030 that is less than about 0.1 percent differ-

8. Barron and others (2019) provide an accessible summary of the study.

ent than in the no-carbon-tax reference scenario (Barron, Halfstead, and Morris 2019).

Metcalf nicely summarizes the state of understanding of the likely incidence of a carbon tax across income classes, noting the importance of both price and income changes (Goulder and others 2019). Even though the policy is now thought to be distributionally neutral or slightly progressive, one may be concerned about any net cost to poor households—even if, as a share of income, it is smaller than the burden on higher-income households. Other research suggests that, at least in the early years, 15 percent or so of the revenue targeted to the lowest three income deciles can hold them harmless on average (Mathur and Morris 2014).

An important limitation of current models is their inability to project longer-term reductions from nascent technologies.[9] For example, few computable general equilibrium models disaggregate the transportation sector to account separately for electric vehicles (EVs), in part owing to the small share of EVs in the current vehicle fleet. Further, the environmental benefits of EVs depend importantly on the emissions intensity of the local power grid, and many computable general equilibrium models do not spatially disaggregate the power sector. This means that models underpredict the emissions reductions available from the transportation sector, but it is unclear by how much. It will remain important to update models as technology evolves and, in the meantime, apply humility in interpreting projections past the next decade or so.

CONCLUSION The strong consensus among economists in favor of taxing carbon rests on a solid base of peer-reviewed research. Ample evidence suggests that a well-designed excise can be environmentally effective, administrable, economically efficient, and distributionally fair. Where experts disagree is largely around the details of the policy, and differences arise primarily over views as to which approaches are most likely to be politically appealing or durable. Economists need not be doctrinaire about whether the policy is revenue neutral or progressive across the entire income distribution, as long as the policy is adopted, remains durable through inevitable business and political cycles, and leverages to the extent possible additional abatement abroad. Additional research is needed to offer ways to protect low-income and coal-reliant households,

9. A more complete discussion of the benefits and limitations of modeling is given by Barron, Hafstead, and Morris (2018).

optimally revise the tax over time, amend existing regulatory programs, address emissions outside the taxed sources, and cost-effectively induce innovation. Although some academic economic departments may view such research as excessively policy oriented, the profession should broaden its taste to value solutions to one of the most critical challenges facing humanity.

REFERENCES FOR THE MORRIS COMMENT

Australian Department of the Environment and Energy. 2018. "Quarterly Update of Australia's National Greenhouse Gas Inventory: September 2018." https://www.environment.gov.au/system/files/resources/4391288e-fc2b-477d-9f0b-99a01363e534/files/nggi-quarterly-update-sept-2018.pdf.

Barron, Alexander R., Allen A. Fawcett, Marc A. C. Hafstead, James McFarland, and Adele C. Morris. 2018. "Policy Insights from the EMF 32 Study on U.S. Carbon Tax Scenarios." *Climate Change Economics* 9, no. 1: 1–47. doi: https://doi.org/10.1142/S201000781840002X.

Barron, Alexander R., Marc A. C. Hafstead, and Adele Morris. 2019. *Policy Insights from Comparing Carbon Pricing Modeling Scenarios*. Washington: Brookings. https://www.brookings.edu/research/policy-insights-from-comparing-carbon-pricing-modeling-scenarios/.

Brooks, Susanne. 2016. "Ensuring Environmental Outcomes from a Carbon Tax." Blog post, Environmental Defense Fund. http://blogs.edf.org/markets/2016/11/03/ensuring-environmental-outcomes-from-a-carbon-tax/.

Center for Climate and Energy Solutions. 2019. "U.S. State Climate Action Plans." https://www.c2es.org/document/climate-action-plans/.

Climate Leadership Council. 2019. "Economists' Statement on Carbon Dividends." Op-ed, *Wall Street Journal*, January 17. https://www.clcouncil.org/economists-statement/.

Crowley, Kate. 2017. "Up and Down with Climate Politics 2013–2016: The Repeal of Carbon Pricing in Australia." *WIREs: Climate Change* 8, no. 3

CRS (U.S. Congressional Research Service). 2019. *Attaching a Price to Greenhouse Gas Emissions with a Carbon Tax or Emissions Fee: Considerations and Potential Impacts*. Report R45625. Washington: Congressional Research Service. https://fas.org/sgp/crs/misc/R45625.pdf.

Diaz, Delavane, and Klaus Keller. 2016. "A Potential Disintegration of the West Antarctic Ice Sheet: Implications for Economic Analyses of Climate Policy." *American Economic Review* 106, no. 5: 607–11.

EPA (U.S. Environmental Protection Agency). 2018. "Regulatory Impact Analysis for the Proposed Emission Guidelines for Greenhouse Gas Emissions from Existing Electric Utility Generating Units; Revisions to Emission Guideline Implementing Regulations; Revisions to New Source Review Program." https://www.epa.gov/sites/production/files/2018-08/documents/utilities_ria_proposed_ace_2018-08.pdf.

Goulder, Lawrence H., Marc A. C. Hafstead, Gyu Rim Kim, and Xianling Long. 2019. "Impacts of a Carbon Tax across U.S. Household Income Groups: What Are the Equity-Efficiency Trade-Offs?" *Journal of Public Economics* 175: 44–64.

Greenstone, Michael, and Ishan Nath. 2019. "Do Renewable Portfolio Standards Deliver?" Working Paper 2019-62. Chicago: Energy Policy Institute at the University of Chicago.

Hafstead, Marc A. C. 2018. *Considering a Carbon Tax-Gasoline Tax Swap*. Washington: Resources for the Future. https://www.rff.org/publications/issue-briefs/considering-a-carbon-taxgasoline-tax-swap-projected-energy-related-us-co2-emissions-reductions-under-the-market-choice-act/.

IPCC (Intergovernmental Panel on Climate Change). 2018. *Global Warming of 1.5°C*. Geneva: IPCC. http://www.ipcc.ch/report/sr15/.

Mathur, Aparna, and Adele C. Morris. 2014. "Distributional Effects of a Carbon Tax in Broader U.S. Fiscal Reform." *Energy Policy* 66: 326–34.

McFarland, James R., Allen A. Fawcett, Adele C. Morris, John M. Reilly, and Peter J. Wilcoxen. 2018. "Overview of the EMF 32 Study on U.S. Carbon Tax Scenarios." *Climate Change Economics* 9, no. 1: 1–37.

McKibbin, Warwick, Adele C. Morris, and Peter Wilcoxen. 2014. "A Proposal to Integrate Price Mechanisms into International Climate Negotiations." *Asia and the Pacific Policy Studies* 1, no. 3: 600–608.

Morris, Adele C. 2018. *Making Border Carbon Adjustments Work in Law and Practice*. Washington: Brookings-Urban Tax Policy Center. https://www.brookings.edu/research/making-border-carbon-adjustments-work-in-law-and-practice/.

Morris, Adele C., and Samantha Gross. 2018. "Why the Climate Challenge Needs Congressional Action." Blog, Brookings. https://www.brookings.edu/blog/planetpolicy/2018/04/21/why-the-climate-challenge-needs-congressional-action/.

Murray, Brian C., William A. Pizer, and Christina Reichert. 2017. "Increasing Emissions Certainty under a Carbon Tax." *Harvard Environmental Law Review* 41: 14–27.

Rothman, Dale S., Bas Amelung, and Philippe Polomé. 2003. *Estimating Non-Market Impacts of Climate Change and Climate Policy*. Paris: Organization for Economic Cooperation and Development. https://www.oecd.org/env/cc/2483779.pdf.

USGCRP (U.S. Global Change Research Program). 2017. *Climate Science Special Report: Fourth National Climate Assessment, Volume I*. Washington: USGCRP. https://science2017.globalchange.gov/.

U.S. Interagency Working Group on the Social Cost of Greenhouse Gases. 2016. "Technical Support Document: Technical update of the Social Cost of Carbon for Regulatory Impact Analysis under Executive Order 12866." https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/scc_tsd_final_clean_8_26_16.pdf.

World Bank. 2019. "Launch of the Finance Ministers' Coalition for Climate Action." https://www.cape4financeministry.org/events/launch-finance-ministers-coalition-climate-action.

GENERAL DISCUSSION Warwick McKibbin began by expressing appreciation for the paper and its focus on a carbon tax. However, there were a couple issues where he thought the paper could be improved. The key point is that substantial credibility in the future carbon price is necessary for a low-cost abatement option. He wondered if it is possible to simultaneously have in place a carbon tax that will increase at a constant rate, and a credible policy. It is really important to create political constituencies to support the policies. He noted that though the paper looks at examples of places where carbon taxes have been implemented and have survived, one should also look at cases where carbon taxes were implemented and failed in order learn the lessons for the design on carbon pricing policy.

McKibbin cited Australia as a good example of a large policy design failure; in 2011, Australia introduced its Clean Energy Act, which consisted of a carbon tax commencing from July 1, 2012, starting at $AU23 a ton, rising over the next three years. However, in the design, Australia made the mistake of building into the legislation a switch to an Emissions Trading System linked to the European Union's Emissions Trading System, to commence on July 1, 2015. This seemed like a good idea at the point of implementation because it should have been a continuation of the carbon price for Australia. However, after the carbon tax in Australia was implemented, the carbon price in Europe collapsed. This resulted in a rising carbon price in Australia for three years, followed by an expected collapsed carbon price, which created costs and little benefits. This led to a massive backlash: because Australia was a carbon-intensive economy, there were many vested interests to fight the carbon tax policy.

McKibbin noted that based on lessons learned from cases like Australia's, a hybrid approach, as touched upon by Metcalf, is worth greater consideration. The key issues of the short-term cost of carbon and the long-term carbon target need to be addressed in the policy design. This can be done by committing to a long-term target and creating emission permits for the entire period into the distant future. These permits would increase in value over time as the number of permits diminished with the target. These permits could be allocated to fossil-fuel-intensive industry, and to voters as compensation for the additional costs of the carbon price. The value of the carbon assets created would be more valuable than the short-term cost to industry and households at the beginning of the program. After allocation of the long-term permits, the balance sheets of corporations would consist of long-term carbon permits and their existing carbon-intensive assets. This mix of assets has the potential to change the behavior of these corporations, which would then have the ability to sell their carbon permits and generate revenue to finance changes in carbon intensity. Hence, a political constituency in support of carbon policy would be created. This first step is a conventional cap-and-trade system with a long-term dimension.

In the shorter term, McKibbin recommended the creation of a central bank of carbon, whose role would be to sell carbon permits in a given year at a fixed price. This would create a kinked supply curve in the short term, with an exact price of carbon as fixed by the central bank of carbon. This would be the equivalent of a short-term carbon tax, except that the revenue from the tax would go mostly to the owners of the long-term permits in the system, and the central bank of carbon (or the government) would get a small amount of revenue each year. This would create long-term credibility and a futures market in carbon pricing. And this would also not be contingent on the government in power, because there would be vested interests of firms and households holding long-term carbon permits, which would want the policy to survive. Hence, McKibbin concluded, a hybrid-built political constituency would more likely generate a sustained carbon price over time. He stated that though he supported a pure carbon tax in theory, he found it politically vulnerable to a change in central government administrations. This lack of credibility, he thought, is the biggest problem. He argued that it is desirable to bind the hands of future governments so that climate policies do not disappear with election cycles.

Justin Wolfers disagreed with Metcalf's comment that a carbon tax is useful only if it decreases emissions. If a carbon tax does not decrease emissions, it is inelastic, and one would rather tax inelastic factors than tax the labor supply or job creators. Hence, Wolfers observed, carbon taxes are great either way—if they work, then environmentalists will be happy; and if they do not work, then public finance economists will be happy.

Wolfers also observed that Metcalf ran many regressions where the left-hand-side variable was GDP. He thought that trying to measure the GDP effect of a carbon tax is a poorly framed question. All economists understand that a deep problem with GDP is that it fails to price environmental resources, and so it is not the right thing to look at. An alternative is to look at employment effects or including an environmental satellite account in GDP to ensure that things are priced properly. Otherwise, one would risk giving sharp answers to horrible questions.

Finally, Wolfers pointed to a report by the Initiative on Global Markets' Economic Experts Panel, Steven Kaplan's group at the University of

Chicago, which surveys economists on various questions.[1] In 2011, this group asked economists if a tax on the carbon content of fuels would be a less expensive way to reduce carbon emissions than would a collection of policies such as "corporate average fuel economy" requirements for automobiles.[2] The response was not unanimous, with exactly one person disagreeing—Edward Lazear, whose response had nothing to do with economics and was more a judgment about political economy. Lazear's response was that "the magnitude of this problem is so great that no sufficient carbon tax is feasible worldwide." Hence, Wolfers concluded, it is safe to assume that the economics profession is completely on board with carbon taxes.

Steven Davis noted that Lawrence Goulder, in his comment on the paper, briefly touched upon the interaction between carbon taxes and other taxes on factor inputs. Davis stated that it is also worth asking how carbon taxes would interact with existing regulations, in particular whether they would accentuate distortions associated with existing inefficient regulatory structures designed to control carbon.

Second, Davis was struck by Adele Morris's presentation showing an order-of-magnitude decline in the value of benefits attributed to carbon abatement in the transition from the Obama administration to the Trump administration. He observed that this points to a larger institutional problem in the way regulatory processes work, such that there are insufficient checks on some matters that are technocratic in character and involve scientific judgments. Davis stated that this process needs discipline, and he recommended it as a topic of consideration for future research.

Alan Viard thanked Metcalf and the commenters for an excellent discussion of carbon taxation, particularly the political issues. He addressed what he considered to be a conceptually important point: the second-best level of the carbon tax relative to the social cost of carbon. He noted that Metcalf suggested that the carbon tax should be scaled back to account for excess burden, based on results from a model in which all individuals are identical. Viard thought that this model, in which the government has no distributional reasons to use commodity or income taxes, was not a good place to start the analysis.

1. "IGM Economic Experts Panel," IGM Forum, University of Chicago–Booth School of Business, 2019, http://www.igmchicago.org/igm-economic-experts-panel.

2. "Carbon Tax Survey," IGM Forum, University of Chicago–Booth School of Business, 2011, http://www.igmchicago.org/surveys/carbon-tax.

Viard recommended starting from a model in which the government faces a trade-off between efficiency and distribution and imposes income and commodity taxes to reduce economic inequality. Viard noted that researchers using this framework have found that the second-best value of a carbon tax can roughly equal its first-best value. Viard agreed with Metcalf that the assumptions for the second-best tax to exactly equal the first-best tax were stringent, but thought that they were a better place to start the analysis than the assumption that there is no inequality in the economy. He noted that, if the existing tax system has design flaws, then the model must be modified to account for the interaction of the carbon tax (in conjunction with the use of carbon tax revenue) with those preexisting flaws. The carbon tax should be scaled back if it reinforces the design flaws and should be scaled up if it alleviates them.

Donald Marron thought that the paper and discussion were great and highlighted some issues. First, he wondered how big a carbon tax ought to be. This is a central issue that is quite hard to answer. One way to address it is by setting the carbon tax equal to the social cost of carbon and the externalities. However, it is hard to quantify the social cost of carbon. Marron commended the Obama administration's efforts in attempting to estimate this social cost. Modeling assumptions addressing climate change, economics, and behavioral responses produce a broad range of plausible estimates of this cost. As a possible solution, Marron recommended reverse engineering carbon taxes by calibrating the tax rates with target levels of emissions.

Second, Marron inquired of the author and commenters regarding their goals. He noted Adele Morris's remarks stating that cutting down emissions to a certain level under the Paris Agreement is one of the goals. Conversely, Marron pointed out, debates about issues like the Green New Deal's target of achieving net zero emissions, and not just cutting emissions, which are two different stories. A carbon tax is an effective tool for achieving a cut in emissions; however, it cannot be the primary tool if the goal is to get to net zero emissions. Although a carbon tax of $50 per ton has the potential to achieve the former goal, the latter goal would need a tax of hundreds of dollars per ton. Marron noted that for a simpler understanding, the price of a carbon tax can be multiplied by 0.1 to figure out the cost per gallon of gasoline as a first approximation. Hence, a tax of $50 per ton is roughly 50 cents a gallon, and a tax of $300 per ton would be $3 a gallon. Finally, Marron noted that there also ought to be subsidies for carbon capture. He observed that Morris touched upon this in her comment, and Marron wondered what the paper's author and others thought about it. He recommended carbon capture subsidization as a potential use of the revenue from carbon taxes.

Richard Cooper observed that British Columbia has lots of hydropower, and so the carbon tax implemented there was essentially a transportation tax. He stated that based on the information that he had, which could be outdated, the fishing industry in British Columbia, which is a big industry with plenty of employment, was excluded from the carbon tax, as well as cruise ships. He asked Metcalf about the accuracy of this information and if this exclusion had since been rectified.

Cooper also stated that in his opinion, a cap-and-trade system cannot be made to work worldwide. Although Europe, the United States, and California can make it work, this would not hold on a global level. This strongly leads Cooper to favor a carbon tax, particularly an international carbon tax, such that it is a common tax whereby the revenue is collected by each country and there is no cross-border revenue sharing. He noted that in addition to avoiding climate change, this will also preserve the world's open trading system, which is important. He concluded by affirming that he was strongly in favor of a uniform or roughly uniform carbon tax worldwide. It does not have to be universal, but a tax spanning over two dozen major countries would be a good starting point.

Jason Furman thought that one of the peculiarities of the carbon tax literature is that individuals have been contributing to it for a long time, developing arguments for a carbon tax. However, there has been no policy progress, especially in the United States. Though in his paper Metcalf did consider cases where there has been progress in carbon taxation, it is important to consider the politics associated with it. Furman acknowledged that Metcalf and the commenters have indeed attempted to address this. He wondered if imperfect action is a substitute or a complement for better action. The Clean Power Plan under the Obama administration, for example, is an open question. The less effectively that the Clean Power Plan was designed, the costlier it would have been for the power sector. A cost-benefit analysis comparison between the Clean Power Plan and a carbon tax would probably have favored the Clean Power Plan.

Furman added that the political economy of a carbon tax is distinct from its economics, and more analysis along these lines would be helpful. He noted that he had seen a lack of research addressing regulatory swaps—which, he acknowledged, is relatively harder to study because it involves taking into account multiple regulations. However, he thought that a more refined understanding of swaps could play a role in improving the political economy of carbon taxation. Finally, he commented on Goulder's point that economists' goal should be maximizing GDP. Furman did not think that politicians had the same goal, and he noted that there is a wide range of social functions, such as maximizing mean income, that could be viable alternatives to maximizing GDP.

James Stock acknowledged the importance of Furman's and Wolfers's points about the economic effects of carbon taxes. Stock noted that in the earlier stages of this project, he was hopeful that Metcalf would be able to come up with credible panel estimates that looked at the different experiences of carbon taxes across countries. However, in hindsight, that seemed like a really tough task because all countries have different experiences. Though Australia's carbon tax only lasted for a while, Sweden applied it only to the transportation sector. Hence, future researchers will need to be aware of the difficulties of determining the empirical evidence on the overall economic effects of carbon taxes. Nonetheless, Stock thought there is potential for further discussion of cases like that of Sweden, although their data sets are messy.

Stock also noted that though much of the discussion has been focused on climate, there are also other co-benefits from the elimination of fossil fuels and action regarding ozone effects. Finally, Stock observed that at the time of this discussion, about 3,300 economists had signed a letter supporting carbon taxes, so it is clear that the economics profession supports them.[3] However, one should not consider their job done once a carbon tax has been passed. In fact, Stock thought that in the shorter term, carbon taxation is the less important policy, and the most important policy would focus on driving down the costs of green alternatives. There have been drops in the prices of wind power, solar photovoltaics, and electric vehicles. These drops have been driven not only by research and development but also by production subsidies, about which economists are typically squeamish. A "learning by doing" approach helps these technologies achieve economies of scale and consequently become more preferable in the market. This shift of the marginal abatement cost curve for these technologies is essential. Hence, Stock concluded, in addition to carbon taxation, it is also important to pursue policies that push these alternative technologies.

Gilbert Metcalf thanked everyone for their comments and thoughts, and noted that he appreciated the points made by the commenters. Regarding Goulder's point about the urgency of the issue, Metcalf noted that there are

3. "Economists' Statement on Carbon Dividends," Climate Leadership Council, January 17, 2019, https://www.clcouncil.org/economists-statement/.

different dimensions on which one can assess a carbon tax—including its efficiency, equity, ease of administration, and political viability, which are all important factors to consider. On the issue of setting the tax rate relative to marginal social damage, Metcalf admitted that there is little consensus on the value of this damage. Accordingly, he agreed with Marron's point of setting a tax rate in terms of domestic emissions reductions. Metcalf wondered what the suitable tax rate for the United States would need to be to bring developing countries on board in international negotiations, where cutting emissions is especially important.

On Cooper's recommendation of a harmonized price, Metcalf noted that Martin Weitzman had also been arguing along the same lines. Though Metcalf was not sure if the international community was ready for a harmonized price, he thought that it was a good idea. Metcalf also answered Cooper's question regarding the fishing industry in British Columbia by confirming that it is indeed excluded from the policy. Regarding Davis's and Wolfers's points about incorporating the benefit of other regulations into the cost of the carbon tax, Metcalf appreciated their cogency and acknowledged that he had not incorporated this analysis in his paper; nor had he come across a good assessment of this question.

Regarding Goulder's observation about a clean energy standard, Metcalf, acknowledging Morris's comments highlighting electricity as a third of the problem, stated that a clean energy standard is not comprehensive and that its implementation would need to be combined with an electrification of the vehicle fleet. He admitted that he was not a big fan of a clean energy standard. And he noted that the big unknowns in any of these analyses are the new technologies that are expected to come along, which reflects the important role of induced innovation. Referring to Stock's comments, Metcalf observed that even modest policies have led to dramatic reductions in the cost of batteries, and in wind and solar resources, so it is important to factor technologies into economic models. Also, this situation is a reflection of the fact that all the cost estimates in the existing models are, in fact, at the upper bounds. Finally, Metcalf noted that negative emission technologies are going to be extremely important and need to be subsidized. Carbon capture and storage from power plants burning coal should receive tax credits equal to the carbon tax rate.