Janice Eberly and James H. Stock, Editors

Brookings Papers
ON ECONOMIC ACTIVITY

FALL 2021

POPP, VONA, MARIN, and CHEN
The Employment Impact of a Green Fiscal Push: Evidence from the American Recovery and Reinvestment Act

BUCKMAN, CHOI, DALY, and SEITELMAN
The Economic Gains from Equity

CHERRY, JIANG, MATVOS, PISKORSKI, and SERU
Government and Private Household Debt Relief during COVID-19

RENNERT, PREST, PIZER, NEWELL, ANTHOFF, KINGDON, RENNELS, COOKE, RAFTERY, ŠEVČÍKOVÁ, and ERRICKSON
The Social Cost of Carbon: Advances in Long-Term Probabilistic Projections of Population, GDP, Emissions, and Discount Rates

REIS
Losing the Inflation Anchor

AUTOR, DORN, and HANSON
On the Persistence of the China Shock

Brookings Papers ON ECONOMIC ACTIVITY

BPEA
FALL
2021

BROOKINGS

BROOKINGS

www.brookings.edu

ISBN 978-0-8157-3973-9

15985-00a_Cover.indd 1 5/12/22 12:11 PM

Brookings Papers
ON ECONOMIC ACTIVITY

FALL 2021

JANICE EBERLY
JAMES H. STOCK
Editors

BROOKINGS INSTITUTION PRESS
Washington, D.C.

15985-00b_FM-2ndPgs.indd 1 5/12/22 12:11 PM

Copyright © 2022 by
THE BROOKINGS INSTITUTION
1775 Massachusetts Avenue, N.W., Washington, D.C. 20036

ISSN 0007-2303
ISBN: Print 978-0-8157-3973-9; ebook: 978-0-8157-3974-6

Authorization to photocopy items for internal or personal use or the internal or personal use of specific clients is granted by the Brookings Institution for libraries and other users registered with the Copyright Clearance Center Transactional Reporting Service, provided that the basic fee is paid to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For more information, please contact CCC at (508) 750-8400 or visit www.copyright.com. This authorization does not extend to other kinds of copying, such as copying for general distribution, for creating new collective works, or for sale. Specific written permission for such copying must be obtained from the Permissions Department, Brookings Institution Press, 1775 Massachusetts Avenue, N.W., Washington, D.C. 20036; e-mail permissions@brookings.edu.

Brookings Papers
ON ECONOMIC ACTIVITY

FALL 2021

DAVID POPP, FRANCESCO VONA, GIOVANNI MARIN, and ZIQIAO CHEN
The Employment Impact of a Green Fiscal Push: Evidence from the American Recovery and Reinvestment Act 1
Comment by Gabriel Chodorow-Reich 50
Comment by Valerie A. Ramey 57
General Discussion 65

SHELBY R. BUCKMAN, LAURA Y. CHOI, MARY C. DALY, and LILY M. SEITELMAN
The Economic Gains from Equity 71
Comment by Nicole Fortin 112
Comment by Erik Hurst 130
General Discussion 135

SUSAN CHERRY, ERICA JIANG, GREGOR MATVOS, TOMASZ PISKORSKI, and AMIT SERU
Government and Private Household Debt Relief during COVID-19 141
Comment by Pascal Noel 200
Comment by Susan Wachter 210
General Discussion 218

KEVIN RENNERT, BRIAN C. PREST, WILLIAM A. PIZER, RICHARD G. NEWELL, DAVID ANTHOFF, CORA KINGDON, LISA RENNELS, ROGER COOKE, ADRIAN E. RAFTERY, HANA ŠEVČÍKOVÁ, and FRANK ERRICKSON
The Social Cost of Carbon: Advances in Long-Term Probabilistic Projections of Population, GDP, Emissions, and Discount Rates 223
Comment by Michael Greenstone 276
Comment by Mar Reguant 295
General Discussion 303

15985-00b_FM-2ndPgs.indd 3 5/12/22 12:11 PM

RICARDO REIS
Losing the Inflation Anchor 307
Comment by Alan S. Blinder 362
Comment by Yuriy Gorodnichenko 369
General Discussion 377

DAVID AUTOR, DAVID DORN, and GORDON HANSON
On the Persistence of the China Shock 381
Comment by Marianne Bertrand 448
Comment by Edward L. Glaeser 457
General Discussion 473

15985-00b_FM-2ndPgs.indd 4 5/12/22 12:11 PM

PURPOSE

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

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

The views expressed by the authors, discussants, and conference participants in *BPEA* are strictly those of the authors, discussants, and conference participants and not of the Brookings Institution. As an independent think tank, the Brookings Institution does not take institutional positions on any issue. The recordings of the Fall 2021 *BPEA* conference can be accessed at https://www.brookings.edu/events/bpea-fall-2021-conference/.

CALL FOR PAPERS

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

ACCESSING THE JOURNAL

All past editions of the *BPEA*, including versions of the figures in color—along with appendix materials, data, and programs used to generate results—are made freely available for download at www.brookings.edu/bpea/search. To purchase 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 *BPEA* are also available through JSTOR (www.jstor.org).

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

Henry Aaron *Brookings Institution*
Stephanie Aaronson *Brookings Institution*
Joseph Abadi *Federal Reserve Bank of Philadelphia*
Katharine Abraham *University of Maryland*
George Akerlof *Georgetown University*
Philippe Andrade *Federal Reserve Bank of Boston*
David Anthoff *University of California, Berkeley*
Alan Auerbach *University of California, Berkeley*
David Autor *Massachusetts Institute of Technology*
Agustin Barboza *Universidad de Montevideo*
Susanto Basu *Boston College*
Michael Bauer *Universität Hamburg*
Joe Beaulieu *Brevan Howard*
Alessandro Bellocchi *Università di Urbino Carlo Bo*
Ben Bernanke *Brookings Institution*
Marianne Bertrand *University of Chicago Booth School of Business*
Sandra Black *Columbia University*
Olivier Blanchard *Peterson Institute for International Economics*
Alan S. Blinder *Princeton University*
Barry Bosworth *Brookings Institution*
Heather Boushey *Council of Economic Advisers*
William Brainard *Yale University*
Steven Braun *Council of Economic Advisers*
Danny Bressler *Council of Economic Advisers*
Jesse Bricker *Federal Reserve Board of Governors*
Markus Brunnermeier *Princeton University*
Ralph C. Bryant *Brookings Institution*
Elaine Buckberg *General Motors*
Shelby R. Buckman *Stanford University*
David Byrne *Federal Reserve Board of Governors*
Ricardo Caballero *Massachusetts Institute of Technology*
Dario Cardamone *University of Notre Dame*
Ziqiao Chen *Syracuse University*
Susan Cherry *Stanford University*
Gabriel Chodorow-Reich *Harvard University*
Laura Y. Choi *Federal Reserve Bank of San Francisco*
Bradley Clark *Council of Economic Advisers*
Gerald Cohen *University of North Carolina Kenan-Flagler Business School*
Susan Collins *University of Michigan*
Charles Collyns *International Monetary Fund*
Roger Cooke *Resources for the Future*
Mary C. Daly *Federal Reserve Bank of San Francisco*
Steven Davis *University of Chicago*
Brad DeLong *University of California, Berkeley*
Jane Dokko *Federal Reserve Bank of Chicago*

David Dorn *University of Zurich*
John Driscoll *Federal Reserve Board of Governors*
Karen Dynan *Harvard University*
Janice Eberly *Northwestern University*
Wendy Edelberg *Brookings Institution*
Barry Eichengreen *University of California, Berkeley*
William English *Yale University*
Frank Errickson *Princeton University*
Adam Falk *Alfred P. Sloan Foundation*
Bruce Fallick *Federal Reserve Bank of Cleveland*
Raquel Fernandez *New York University*
Alessandra Fogli *Federal Reserve Bank of Minneapolis*
Kristin Forbes *MIT Sloan School of Management*
Nicole Fortin *University of British Columbia*
Benjamin Friedman *Harvard University*
Jason Furman *Harvard University*
Amy Ganz *Aspen Institute*
Edward L. Glaeser *Harvard University*
Austan Goolsbee *University of Chicago Booth School of Business*
Robert Gordon *Northwestern University*
Tracy Gordon *Urban-Brookings Tax Policy Center*
Egor Gornostay *Peterson Institute for International Economics*
Yuriy Gorodnichenko *University of California, Berkeley*
Lawrence Goulder *Stanford University*
Michael Greenstone *University of Chicago*
John Grigsby *Princeton University*
Erica Groshen *Cornell University*
Robert Hall *Stanford University*
John Haltiwanger *University of Maryland*
Gordon Hanson *Harvard Kennedy School*
Susan Helper *Council of Economic Advisers*
Peter Blair Henry *New York University Stern School of Business*
Bart Hobijn *Arizona State University, Tempe*
Harry Holzer *Georgetown University*
Caroline Hoxby *Stanford University*
Erik Hurst *University of Chicago Booth School of Business*
Ethan Ilzetzki *London School of Economics*
Douglas Irwin *Dartmouth College*
Erica Jiang *University of Southern California*
Zeynal Karaca *Agency for Healthcare Research and Quality*
Amit Khandelwal *Columbia Business School*
Anupam Khanna *formerly of NASSCOM*
Michael Kiley *Federal Reserve Board of Governors*
Cora Kingdon *University of California, Berkeley*
Donald Kohn *Brookings Institution*
Jacob Krimmel *Federal Reserve Board of Governors*
Byron Lutz *Federal Reserve Board of Governors*
Dean Maki *Point72 Asset Management*
Greg Mankiw *Harvard University*
Giovanni Marin *University of Urbino Carlo Bo*
Gregor Matvos *Northwestern University*
Mark Mazur *Department of the Treasury*
Laurence Meyer *LHMeyer/Monetary Policy Analytics*
Evan Michelson *Alfred P. Sloan Foundation*

Gian Maria Milesi-Ferretti *Brookings Institution*
Jeffrey Miron *Harvard University*
Frederic Mishkin *Columbia University*
Robert Moffitt *Johns Hopkins University*
Adele Morris *Federal Reserve Board of Governors*
Richard G. Newell *Resources for the Future*
Pascal Noel *University of Chicago Booth School of Business*
Maurice Obstfeld *University of California, Berkeley*
Giovanni Olivei *Federal Reserve Bank of Boston*
Abaigeal O'Shea *Council of Economic Advisers*
Jonathan Parker *Massachusetts Institute of Technology*
George Perry *Brookings Institution*
Thomas Philippon *New York University Stern School of Business*
Tomasz Piskorski *Columbia University*
William A. Pizer *Resources for the Future*
David Popp *Syracuse University*
Brian C. Prest *Resources for the Future*
Adrian E. Raftery *University of Washington*
Valerie A. Ramey *University of California, San Diego*
Tristan Reed *World Bank Group*
Mar Reguant *Northwestern University*
Ricardo Reis *London School of Economics*
Lisa Rennels *University of California, Berkeley*
Kevin Rennert *Resources for the Future*
Kevin Rinz *Council of Economic Advisers*
Liliana Rojas-Suarez *Center for Global Development*
David Romer *University of California, Berkeley*
Robert Rosener *Morgan Stanley*
Glenn Rudebusch *Federal Reserve Bank of San Francisco*
Claudia Sahm *Jain Family Institute*
Edgar J. Sanchez Carrera *University of Urbino Carlo Bo*
Brian Scholl *Securities and Exchange Commission*
Lily M. Seitelman *Boston University*
Amit Seru *Stanford University*
Hana Ševčíková *University of Washington*
Jay Shambaugh *George Washington University*
Matthew Shapiro *University of Michigan*
Louise Sheiner *Brookings Institution*
Chris Sims *Princeton University*
Tara Sinclair *George Washington University*
Mark Steinmeyer *Smith Richardson Foundation*
James H. Stock *Harvard University*
Michael Strain *American Enterprise Institute*
Phillip Swagel *Congressional Budget Office*
Daniel Tarullo *Harvard Law School*
Robert Tetlow *Federal Reserve Board of Governors*
Angel Ubide *Citadel*
John Van Reenen *London School of Economics*
Stan Veuger *American Enterprise Institute*
Alan Viard *American Enterprise Institute*
Annette Vissing-Jorgensen *Federal Reserve Board of Governors*
Alice Volz *Federal Reserve Board of Governors*
Francesco Vona *OFCE Sciences-Po and University of Milan*
Susan Wachter *University of Pennsylvania Wharton School*

15985-00b_FM-2ndPgs.indd 8 5/12/22 12:11 PM

William Wascher *Federal Reserve Board of Governors*
Mark Watson *Princeton University*
David Wessel *Brookings Institution*
David Wilcox *Peterson Institute of International Economics*
Jon Willis *Federal Reserve Bank of Atlanta*
Justin Wolfers *University of Michigan*
Luigi Zingales *University of Chicago*
Eric Zitzewitz *Dartmouth College*

Mitchell Barnes *Brookings Institution*
Janina Broker *Brookings Institution*
Haowen Chen *Brookings Institution*
Siobhan Drummond *Brookings Institution*
Kelly Kennedy *Brookings Institution*
Conrad Milhaupt *Brookings Institution*
Nasiha Salwati *Brookings Institution*
David Skidmore *Brookings Institution*

15985-00b_FM-2ndPgs.indd 9 5/12/22 12:11 PM

15985-00b_FM-2ndPgs.indd 10 5/12/22 12:11 PM

DAVID POPP
Syracuse University

GIOVANNI MARIN
University of Urbino Carlo Bo

FRANCESCO VONA
OFCE Sciences-Po and University of Milan

ZIQIAO CHEN
Syracuse University

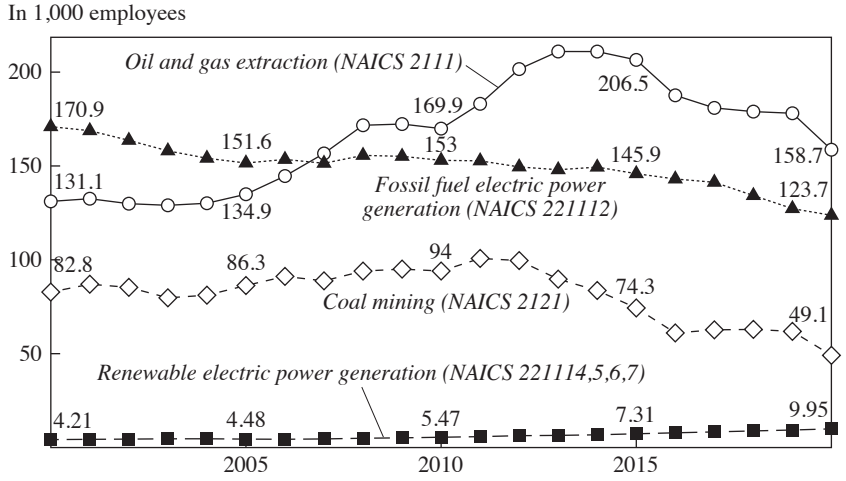
The Employment Impact of a Green Fiscal Push: Evidence from the American Recovery and Reinvestment Act

ABSTRACT Investments in the green economy are used for both environmental goals and fiscal stimulus. The success of these investments depends, at least in part, on whether they create new jobs and whether such jobs are available to workers hurt by a green transition. We evaluate the employment effect of green investments from the American Recovery and Reinvestment Act (ARRA). Most job creation from green ARRA investments is permanent and emerged in the post-ARRA period, but the plausible range of estimates is extremely wide (zero to twenty-five jobs per $1 million). Such large uncertainty on aggregate effects masks substantial heterogeneity across communities. The green stimulus mostly benefited areas with a greater prevalence of preexisting green skills that created 40 percent additional jobs than average communities. New jobs are primarily manual labor and in occupations performing green tasks, especially in renewable energy. However, manual labor wages do not increase. Descriptive evidence suggests that the skill gap between green energy and fossil fuel workers is modest, but green jobs require significantly more training. Because the spatial distribution of skills and jobs matters, using green stimuli can help reshape the economy in the long run but may also exacerbate regional inequities associated with the green energy transition.

Conflict of Interest Disclosure: This project was supported in part through the Smart Prosperity Institute Research Network and its Greening Growth Partnership, which is supported by a Social Sciences and Humanities Research Council of Canada Partnership Grant (no. 895-2017-1018), as well as by Environment and Climate Change Canada's Economics and Environmental Policy Research Network (EEPRN). Other than the aforementioned, the authors did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this article. They are currently not an officer, director, or board member of any organization with an interest in this article.

15985-01a-Popp-4thPgs.indd 1 5/12/22 12:11 PM

Figure 1. US Energy Industry Employment over Time



Source: Quarterly Census on Employment and Wages, Bureau of Labor Statistics (QCEW-BLS).
Note: Figure shows employment (full-time equivalent, FTE) by industry.

There is growing interest in green fiscal stimuli. Investing in the green economy has been identified as a strategic area of intervention as a response both to the climate crisis and to the economic crisis induced by the COVID-19 pandemic (Helm 2020; Agrawala, Dussaux, and Monti 2020). President Biden's original American Jobs Plan, unveiled in March 2021, includes more than $500 billion in green investments such as electric vehicle charging stations, modernizing the electricity grid, and improving climate resilience (White House 2021a). In Europe, the European Commission's European Green Deal, first proposed prior to the COVID-19 pandemic in December 2019, puts a green fiscal stimulus at the center of the European Union's growth strategy to achieve social, economic, and environmental goals.[1]

Among the goals of most green fiscal stimuli is creating new green jobs for workers potentially displaced by a green transition. Figure 1 shows recent US employment trends in energy industries. Employment in coal mining fell from a peak of just over 100,000 in 2011 to around 49,000 in

1. European Commission, "A European Green Deal: Striving to Be the First Climate-Neutral Continent," https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en.

2020. Employment in fossil fuel electric power generation has also fallen. Driven by the boom of the shale gas revolution, employment in oil and gas extraction thrived in the same time period, growing from just over 131,000 workers in 2000 to nearly 211,000 by 2013. However, employment in the sector has fallen each year since. While employment in renewable electric power production has doubled over the last decade, these workers remain a small share of jobs in the energy sector, with around 9,950 workers employed in 2020.

In the long run, the acceleration in renewable energy investments triggered by both a green stimulus package and climate policy more generally will pose a significant threat not only to coal communities but to the prosperity of communities depending heavily on oil and gas. The threats faced by these communities are a barrier to political support for carbon pricing and climate policy in the United States and elsewhere (Tomer, Kane, and George 2021; Weber 2020; Vona 2019). To address these concerns, the American Jobs Plan specifically targets hard-hit mining communities, proposing investments such as plugging orphan oil and gas wells and cleaning abandoned mines to create jobs while improving local environmental conditions. Similarly, the European Green Deal includes €17.5 billion to aid regions and workers most affected by a green energy transition.

More generally, the success of green fiscal stimuli depends, at least in part, on whether these investments create new jobs and whether such jobs are available to workers for whom a green transition has a negative impact. Three sets of questions help inform the potential role of green fiscal stimuli as part of a green energy transition.

First, what does the existing literature say about the effect of environmental policies on employment? Will the transition to clean energy create winners and losers among workers with different sets of skills? As we show in section I, while the net effects on employment may be small, recent studies suggest more nuanced results, with low-skilled manual labor workers bearing the largest burden of energy price increases.

Second, what are the employment options for workers in polluting industries with declining job prospects? What must we know to determine their ability to be reemployed? Labor economics research shows that reallocation costs are proportional to the skill differences between origin and destination jobs. To assess whether the skills of displaced workers are likely to be needed in a green economy, in section II we provide descriptive evidence on the skill similarity between green and brown workers.

Third, to what extent can government investments, such as the American Jobs Plan or European Green Deal, be used to create new jobs? And

to what extent are any resulting job gains heterogeneous across regions and occupations? To answer this question, our paper uses the US experience from the 2009 American Recovery and Reinvestment Act (ARRA) to assess the potential employment impacts of green fiscal stimuli as part of a green transition.

Our paper provides the first rigorous assessment of the employment effects of ARRA's green investments by exploiting the cross-sectional variation of green ARRA spending across regions. The full stimulus package included over $500 billion of direct government spending, and an additional $290 billion in tax reductions (Council of Economic Advisers 2014). We focus on the direct spending targeted at green investments, which constituted approximately 19 percent of all direct government spending in ARRA (online appendix figure A1). Because a large share of green spending was devoted to public investments, green ARRA may have a cumulative effect stretching beyond the stimulus period (Council of Economic Advisers 2013, 2014). We thus differentiate between the short- and long-term effect of green ARRA. Overall, we find that the effect of green ARRA on job creation emerges primarily in the long run. However, the plausible range of long-term effects is very wide (ranging from zero to twenty-five jobs per $1 million spent), as regions receiving more green spending were more resilient to the Great Recession. The timing of green ARRA's impact differs from previous studies of other ARRA investments, which generally find larger short-term effects.

Importantly, the impact of green ARRA becomes much clearer when we explore several dimensions of heterogeneity. First, green ARRA creates more jobs in commuting zones with a greater prevalence of preexisting green skills. Roughly speaking, green ARRA spending creates up to 70 percent more jobs in areas in the top quartile of the green skills distribution than in the median commuting zone. As the presence of green skills in a community is also strongly correlated with the allocation of green ARRA subsidies, our results provide evidence of the green stimulus as a successful example of picking winners.

Second, looking at specific sectors of the economy, we see the potential of a green stimulus to reshape an economy and have important distributional effects. Most new jobs are manual labor positions, with much growth in the green and construction sectors. These new manual labor positions are permanent jobs, suggesting that green spending can help improve labor market conditions for unskilled workers. However, manual labor wages did not increase. Our research suggests that there may be a suitable path for reallocating manual workers displaced by carbon pricing policies in

15985-01a-Popp-4thPgs.indd 4 5/12/22 12:11 PM

energy-intensive and fossil fuel industries into green jobs in construction and waste management. However, these green jobs require substantially more on-the-job training than brown jobs, so reallocation costs will not be negligible. Our analysis indicates that job training investments should be a crucial part of any green (energy) transition seeking to minimize labor market distributional effects.

Finally, we make a number of distinct contributions to the empirical literature on fiscal multipliers by looking specifically at investments in the green economy, which are likely to become increasingly important in the future. Our findings are directly comparable with those of the broader literature estimating the effects of ARRA.[2] But the wide range of plausible values of the job creation effect does not allow us to reach firm conclusions on whether green spending is more effective than other types of spending in creating jobs. In the spirit of recent contributions seeking to isolate the microeconomic mechanisms of the local multiplier (Moretti 2010; Garin 2019; Dupor and McCrory 2018; Auerbach, Gorodnichenko, and Murphy 2020), we study the time profile of the effect, the role of key mediating factors, distributional effects across workers, and some mechanisms through which the green stimuli have an impact on the local economy. Results become much clearer when these additional dimensions are added to the analysis. Thus, our paper presents new evidence on the reshaping effect of a green fiscal push rather than providing a precise estimate of the aggregate effect.

Aggregate effects of a green fiscal stimulus are more difficult to identify than the effect of other types of ARRA spending. For other ARRA spending, endogeneity concerns primarily result from the fact that ARRA spending was focused on areas hardest hit by the Great Recession. To identify the causal effects of ARRA, previous literature isolates the exogenous component of the geographical allocation of the fiscal stimulus, and thus a causal effect, using preexisting formulas to allocate federal funds (Wilson 2012; Chodorow-Reich and others 2012; Nakamura and Steinsson 2014). Identifying the causal effect of the green stimulus presents an additional challenge, as the allocation of green ARRA spending depends on structural characteristics of the economy, such as greater technological expertise in green communities, that are positively correlated with employment growth. While we discuss these issues extensively in section IV, we anticipate that these solutions do not completely remove pre-trends between

2. See Chodorow-Reich (2019) for a survey of this literature.

overall employment growth and green ARRA investments. However, this violation of a parallel trend assumption matters only for total employment. When looking at specific sectors or occupations we find no evidence of pre-trends, providing us with confidence that these results are more credible and easier to interpret.

I. Understanding the Employment Effects of Environmental Policies

Besides contributing to the literature on fiscal stimulus, our research is informed by previous work on the employment effects of various environmental policies. Here we present key findings from this literature and discuss our contributions. The effect of environmental policy on employment is still hotly debated and polarized, with advocates on both sides ignoring or exaggerating the labor market costs and benefits of environmental regulations.[3] Advocates of stronger environmental policies argue that such policies create high-paying "green jobs," while critics point to the job losses in energy-intensive industries and mining activities that they are sure will follow. Previous literature finds that the net effect of environmental policies on employment is small, especially when general equilibrium effects and offsetting mechanisms are accounted for (Morgenstern, Pizer, and Shih 2002; Hafstead and Williams 2018; Metcalf and Stock 2020). Moreover, a larger pool of workers with the skills required to perform green tasks reduces mobility frictions and reallocation costs, thus improving the aggregate effect of environmental policies (Castellanos and Heutel 2019). However, recent studies find more nuanced results when looking at specific sectors or workers, with job losses concentrated in polluting industries and among unskilled workers (Yip 2018; Marin and Vona 2019), while workers with technical and engineering skills experience increased employment (Vona and others 2018).

Addressing effects across industries, Kahn and Mansur (2013) compare employment at the county level for adjoining counties, one of which is in attainment with Clean Air Act air quality standards and one which is not. Counties not in attainment face more stringent air pollution regulations, and using neighboring counties as controls helps control for other factors likely to affect employment. Kahn and Mansur find that non-attainment status leads to job losses in specific industries that are intensive in electricity,

3. See, for instance, the survey in Vona (2019).

labor, and pollution. Examples of such industries include petroleum products, paper, primary metals, and transportation equipment. The effect is equivalent to the job losses that would result from a 33 percent increase in electricity prices in attainment counties.

Yip (2018, 2020) uses a difference-in-differences approach to identify the effect of the British Columbia carbon tax on workers using a rich individual-level data set. His main finding is that the tax disproportionately harms workers with middle and low levels of education, both in terms of increases in unemployment rates and decreases in wages. Marin and Vona (2019) examine the effect of long-term increases in energy prices on the relative demand of coarse occupational groups (managers, professionals, technicians, and manual workers) for EU countries and industrial sectors over the period 1995–2011.[4] In their preferred specification, which controls for endogeneity and intervening factors such as import penetration and investments in information and communication technology, the large historical increase in energy prices explains about 13.5 percent of the concomitant increase in the share of technicians, but just 5 percent of the decline in the share of manual workers. Adverse impacts on manual workers are in general small, but they are of particular concern for the political acceptability of green policies since they amplify the secular deterioration of their labor market conditions driven by automation and globalization (Autor, Levy, and Murnane 2003; Autor, Dorn, and Hanson 2013).

In one of the rare studies directly evaluating transitional costs using individual-level data, Walker (2013) shows that foregone lifelong earnings for workers displaced by the US Clean Air Act are larger for workers who change sector. More generally, labor research shows that reallocation costs are proportional to the skill proximity between origin and destination jobs (Kambourov and Manovskii 2009; Gathmann and Schönberg 2010). Thus, the job creation effect of a green stimulus may depend on the availability of workers with the appropriate skills. The distributional effects are expected to be smaller if displaced workers (e.g., coal miners) possess skills that are important to perform the tasks required by the new green jobs (e.g., solar photovoltaic installers). We shed light on the first issue by letting the effect of green spending vary depending on the green skills available in the

4. Historical variation in energy prices has been used as a proxy to evaluate the effects of carbon prices in previous papers (Aldy and Pizer 2015; Cullen and Mansur 2017; Marin and Vona 2021). The reliability of such a proxy is clearly stronger when it is possible to estimate the long-term effect of energy prices (Marin and Vona 2021).

local economy (Vona and others 2018), while we tackle the second issue by comparing the skill and training requirement of workers in the green and brown occupations.

While much work evaluates the effect of policies imposing a cost of pollution (either through standards or prices) on labor markets, almost no work considers the potential of green subsidies opening up new employment opportunities in the so-called green economy. The only exception is the related paper by Vona, Marin, and Consoli (2019), which uses similar data. Following Moretti (2010), they estimate the additional number of jobs indirectly created in the local economy by a new green job. We extend their work by estimating the direct effect of green subsidies, its time profile, and the heterogeneous effects across workers, sectors, and communities.

The effects of a green fiscal push and of carbon taxation (and energy prices) are similar on potential job losers, as both accelerate job destruction in fossil fuel–intensive sectors, possibly creating inequities across regions, sectors, and workers. Regional effects of greening the economy are a politically sensitive issue, especially in the United States, where the fossil fuel industry provides geographically concentrated jobs and drives the economic growth of many local communities (Tomer, Kane, and George 2021). Fearing losses of income and employment, these local communities are reluctant to support any transition without a clear alternative, which is a factor in low political support for carbon pricing and climate policy overall in the United States (Tomer, Kane, and George 2021; Weber 2020). Thus, advocates for a "just transition" argue that carbon pricing alone is unlikely to succeed for both ethical and practical reasons, and a more comprehensive approach is needed to achieve the equity goals (Konisky and Carley 2021), including consideration of the skills and characteristics of occupations needed for a green economy (Muro and others 2019). Despite the challenges to meet the broad equity goals, there can be much potential for crafting policy solutions that bring to these communities more jobs in the clean energy industry, as the geographic distribution of clean energy resources largely aligns with that of fossil fuel resources in the United States (Tomer, Kane, and George 2021).

II. Evidence on Green Skills and Employment

As our empirical analysis will show, the potential for green investments to create jobs depends, at least in part, on a good match between the skills of workers and the jobs being created. Thus, if green fiscal stimuli are to help smooth the employment transition for fossil fuel workers, it is important

to compare the skills of these workers to those jobs likely created by green investments. Using data from the Department of Labor's Occupational Information Network (O*NET) to construct green general skills (GGS) indexes, as in Vona and others (2018), we compare the characteristics of workers in brown and green sectors of the economy. Green general skills are skills potentially used in all occupations but that are particularly important for green occupations (Vona and others 2018). However, not all jobs using these green skills are "green jobs." Green general skills are also important in occupations such as physicians, mining machine operators, and some transportation workers. The key point is that workers in these jobs have the skills necessary to do the work required of green occupations, making retraining costs associated with job-to-job mobility lower.

In table 1, we report descriptive statistics comparing both low-skilled (LS) and high-skilled (HS) workers in green and brown occupations. We include data on green skills, training requirements, and other characteristics. Both are also compared to a benchmark of all other occupations in a Standard Occupational Classification (SOC) two-digit group containing at least one green or brown occupation. We use definitions of green and brown occupations by Vona and others (2018), which we further divide into energy and non-energy green and brown occupations.[5] Green energy occupations include jobs related to wind and solar energy, as these are expected to be the main beneficiaries of green stimulus investments around the world. The importance of green skills for each task is normalized to vary between zero and one, and is presented for four macro groups of green skills: engineering and technical, operation management, monitoring, and science. Further details on the construction of all data presented here, including the measurement of green skills, are presented in online appendix A2.

The first section of table 1 compares basic descriptive characteristics of each occupation. Brown fossil fuel occupations stand out in terms of hourly wages, especially for HS workers (row 1). Notably, these jobs, which focus on extraction and production of fossil fuels, are more geographically concentrated than comparable jobs (row 2). For low-skilled workers, green renewable occupations have the highest wages, followed by fossil fuel workers. However, the high wages found for renewable energy workers are primarily due to the wages reported for solar energy sales representatives and installation managers. Wages for installers and service technicians are similar to those of comparable fossil fuel energy workers (tables A5 and A6

5. Online appendix A2 summarizes the definitions of green and brown occupations.

15985-01a-Popp-4thPgs.indd 10

Table 1. Characteristics of Different Green and Brown Occupational Groups

	Brown fossil fuel occupations		*Other brown occupations*		*Green renewable energy occupations*		*Other green occupations*		*Benchmark*	
	LS	*HS*	*LS*	*HS*	*LS*	*HS*	*LS*	*HS*	*LS*	*HS*
1. Hourly wage (BLS)	25.80	70.73	22.22	37.54	37.51	49.26	24.35	53.33	19.49	44.52
2. Gini locational coefficient (ACS)	0.73	0.79	0.57	0.66	0.31	0.44	0.38	0.48	0.34	0.47
3. Age (ACS)	38.86	40.82	41.67	40.86	44.08	41.76	41.03	43.29	39.05	42.02
4. Share of male (ACS)	0.97	0.89	0.78	0.62	0.88	0.84	0.90	0.75	0.52	0.50
5. Educational attainment (yrs.) (ACS)	12.06	15.70	12.04	15.36	12.75	15.73	12.31	15.06	12.67	15.13
6. Required months on-the-job training (O*NET)	8.98	22.30	9.36	7.91	12.62	13.60	13.94	16.74	6.32	12.79
7. GGS: engineering and technical (O*NET)	0.43	0.54	0.43	0.39	0.68	0.69	0.44	0.55	0.24	0.27
8. GGS: operation management (O*NET)	0.42	0.71	0.42	0.60	0.55	0.60	0.46	0.63	0.38	0.61
9. GGS: science (O*NET)	0.25	0.41	0.19	0.36	0.26	0.40	0.21	0.37	0.09	0.15
10. GGS: monitoring (O*NET)	0.47	0.58	0.42	0.56	0.52	0.59	0.46	0.57	0.41	0.61

Sources: Bureau of Labor Statistics, Occupational Employment Statistics (BLS); American Community Survey (ACS); Occupational Information Network database (O*NET).

Note: Macro occupational groups are defined in table A4 in online appendix A2. Low-skill (LS) occupations belong to SOC two-digit major groups from 31 to 53; high-skill (HS) occupations belong to SOC two-digit major groups from 11 to 29. The benchmark is defined as all nongreen and nonbrown occupations in SOC two-digit major groups with at least one green or brown occupation. The following SOC two-digit major groups are excluded from the benchmark: 21, 23, 25, 29, 31, 33, 35, 37, 39. Statistics report averages weighted by occupational employment (from BLS) in 2019. Occupational employment for green occupations is further reweighted for the greenness of the occupation, as presented in online appendix A2. The Gini locational coefficient (see Gabe and Abel 2012) is based on data by occupation and commuting zone from the American Community Survey (1 percent sample, 2019). O*NET data refer to the latest release of O*NET (25.3). Green general skills (GGS) scores are based on Vona and others (2018).

5/12/22 12:11 PM

in online appendix A2). Other green occupations are, on average, paid wages greater than in benchmark jobs. While ages of workers are similar across occupations (row 3), a striking difference emerges in terms of gender orientation. The share of males is significantly higher in brown fossil fuel occupations than in other sectors (row 4), suggesting males are more likely to experience negative employment shocks in the transition to a green economy.

Differences in the skill and training requirements represent potential barriers to reemploying brown workers into green jobs. Table 1 illustrates both key similarities and differences. First, the educational requirements of low-skilled brown jobs are slightly lower than both green jobs and the benchmark occupations (row 5). Second, while high-skilled fossil fuel jobs require more months of on-the-job training than other categories (row 6), for low-skilled workers green jobs require substantially more training (at least 12 months) than either the benchmark (6.5 months) or brown occupations (9 months).[6] Third, while green low-skilled occupations require more training, the skills data suggest that green and brown occupations have closer skill sets than the green and benchmark occupations (rows 7–10). This similarity is particularly notable between brown energy jobs and green non-energy jobs. The important difference is that green renewable energy occupations require more engineering and operations management skills than fossil fuel brown occupations. Tables A5 and A6 in the online appendix show that this is true for nearly every possible combination of green energy and brown energy jobs. The role played by the endowment of green skills in the local labor market will be analyzed also in the econometric evaluation of the green ARRA program in section V.

III. The American Recovery and Reinvestment Act

We use data on green investments in the 2009 stimulus package to estimate the impact of green investments on employment. In response to the Great Recession, the American Recovery and Reinvestment Act (ARRA) of 2009, commonly known as the stimulus package, invested over $800 billion in the forms of tax incentives and federal spending programs to stimulate the US economy (Council of Economic Advisers 2014). Through ARRA

6. Notable exceptions shown in table A5 of the online appendix are solar photovoltaic installers and wind turbine service technicians, which require similar levels of training to fossil fuel workers. However, these jobs also require greater engineering and operations management skills, as discussed below.

spending programs, federal agencies partnered with state and local governments and nonprofit and private entities to help "put Americans back to work." Naturally, much of the spending funded projects that provide immediate job opportunities, such as highway construction, or filled state budget shortfalls to bail out school systems and save the jobs of teachers and school staff.

While the primary goal of ARRA was to stimulate macroeconomic growth and provide job opportunities, part of the funds were invested in "transportation, environmental protection, and other infrastructure that will provide long-term economic benefits" (American Recovery and Reinvestment Act of 2009, 2). These include both direct spending intended for immediate job creation, such as Department of Energy (DOE) spending for renewable energy and energy efficiency retrofits and Environmental Protection Agency (EPA) grants for brownfield redevelopment, as well as tax breaks and loan guarantees for renewable energy. Our work focuses on the impact of direct spending intended for job creation, asking both whether these green investments stimulated employment and what types of workers may benefit from a green stimulus.

Among the key principles motivating infrastructure investments in ARRA was that facilitating the transition to an energy-efficient and clean energy economy would lay the foundation for long-term economic growth (US Office of the Vice President 2010). As a result, ARRA included more than $90 billion for clean energy activities, including DOE contracts and grants to support projects such as energy efficiency retrofits, the development of renewable energy resources, public transportation and clean vehicles, and modernizing the electric grid (Aldy 2013). To meet the Obama administration's target of doubling renewable energy generation by 2012, the DOE provided assistance for a large number of projects related to renewable energy. For example, the Clean Energy Technology Center in Massachusetts received $24.8 million to design, construct, and operate a wind turbine blade testing facility (US Department of Energy 2010). Moreover, $3.4 billion in cost-shared grants supported the deployment of smart grid technology, generating more than $4.5 billion of co-investment (Aldy 2013). ARRA funding also supported the expansion of the Weatherization Assistance Program, which supports low-income families for energy efficiency improvements (Fowlie, Greenstone, and Wolfram 2018).

The EPA oversaw most ARRA programs designated for environmental protection. The largest of these programs was $6.4 billion for Clean Water and Drinking Water State Revolving Funds, which are among the programs analyzed in Dupor and McCrory (2018). An additional $600 million was

set aside for the EPA's Superfund program to clean up contaminated sites such as New Bedford Harbor in Massachusetts, to which the EPA allocated $30 million (US Office of the Vice President 2010).[7] Another $200 million was invested in the Leaking Underground Storage Tank Trust Fund for the prevention and cleanup of leakage from underground storage tanks.[8] Other EPA funds were allocated to improvements of infrastructure such as wastewater treatment facilities and diesel emissions reduction (Environmental Protection Agency 2009). Different from other ARRA programs, which were allocated according to statutory formulas based on exogenous factors such as the number of highway lane miles in a state or the youth share of its population (Wilson 2012), much green ARRA funding does not follow the same rules.

III.A. Data on ARRA Awards

Our analysis covers the universe of contracts, grants, and loans awarded under the ARRA between 2009 and 2012. Recipients of ARRA funding are required to submit reports through FederalReporting.gov which include information on the amount of expenses and the description of projects.[9] We retrieved data from FedSpending.org on these records derived from reports submitted by nonfederal entities who received ARRA funding.

In line with most recent evaluations of ARRA (Dupor and Mehkari 2016; Dupor and McCrory 2018), our unit of analysis is the local labor market, that is, the so-called commuting zone (CZ). We aggregate county-level data into 709 commuting zones based on the official CZ definitions from the 2000 Decennial Census. As in Dupor and Mehkari (2016), we exclude 122 commuting zones with less than 25,000 inhabitants in 2008, which represent less than 0.5 percent of the US population and employment. We also drop the commuting zone pertaining to New Orleans, as its employment and population data are heavily influenced by the damages and recovery from Hurricane Katrina. Our primary estimation sample thus contains 587 commuting zones. As the entities known as prime recipients who

7. Information on active and archived Superfund sites is available at United States Environmental Protection Agency, "Search Superfund Site Information," https://cumulis.epa.gov/supercpad/cursites/srchsites.cfm.

8. EPA's Web Archive, EPA Information Related to the American Recovery and Reinvestment Act of 2009, "Leaking Underground Storage Tank Program Implements the Recovery Act," https://archive.epa.gov/oust/eparecovery/web/html/index.html.

9. This website is no longer in use, but archived data are available at National Bureau of Economic Research, "American Recovery and Reinvestment Act (ARRA) Federal Spending Detail 2009–2012," https://data.nber.org/data/ARRA/.

received funding directly from the federal government may make subcontracts to other entities, we use the reported place of performance of prime and subprime recipients to allocate the dollar amount of awards to commuting zones based on the zip code. Our ARRA data are time-invariant and include the total amount awarded between 2009 and 2012. As noted in Wilson (2012), nearly 90 percent of expected ARRA spending had been obligated by 2010.[10]

Nearly all DOE and EPA projects relate to the green economy.[11] Thus, our measure of green ARRA includes all ARRA projects from the DOE and EPA and their subordinate agencies, such as various national laboratories. All other ARRA spending is coded as nongreen ARRA.[12] Table A1 in online appendix A1 provides descriptive data on both green and nongreen ARRA. Overall, the stimulus package included over $61 billion on green investments and almost $262 billion on nongreen investments. Of these green investments, $52 billion come from the DOE, while just $9 billion come from the EPA. Roughly 10 percent of green ARRA spending supported research and development (R&D). A small $228 million supported job training for green occupations.

The mean values of green ARRA and nongreen ARRA per commuting zone in our sample are approximately $103 million and $440 million, respectively. The per capita levels of green ARRA and nongreen ARRA are $260 and $985, respectively, based on population in 2008. We highlight the skewed distribution of both green and nongreen ARRA, as the median commuting zone received only $105 and $819 per capita of green and nongreen ARRA awards.

10. Unlike other evaluations of ARRA, we do not consider the location of vendors when allocating funds. Our goal is to ascertain the effectiveness of green ARRA given the "greenness" of the local economy. If a recipient must use vendors from outside the local commuting zone to satisfy a need of the project due to a lack of qualified suppliers in the commuting zone, the funding has been less effective for stimulating local employment.

11. To verify this, we checked projects with the terms *oil*, *gas*, or *coal* in the description. None of these projects related to discovery of new sources. More commonly, they referenced reducing consumption, clean coal, carbon sequestration, or biofuels as a substitute.

12. In addition to the EPA and DOE, a few other agencies funded investments that were plausibly green. The Department of Labor supported four small job training programs (totaling just $496 million) that focused on energy efficiency and the renewable energy industry. Including these investments as green ARRA does not change our results. While the Department of Housing and Urban Development (HUD) also supported green building retrofits, we did not include these programs in our analysis. These do not fall under a single green program, and thus must be identified manually. In our attempt to label HUD investments as "green," we found that many of the green HUD grants were trivial—for example, installing LED lightbulbs in a building—and should have little to no impact on green employment.

15985-01a-Popp-4thPgs.indd 14 5/12/22 12:11 PM

Figures A2, A3, and A4 in online appendix A1 illustrate the geographic distribution of green ARRA and nongreen ARRA. We do not observe any apparent, systematic patterns across geographic areas, as both areas receiving high per capita amounts (figures A2 and A3) and areas receiving large shares of green stimulus (figure A4) are spread throughout the country (see table A2 for a list of commuting zones that received the largest ARRA awards). Large infrastructure projects play an important role in communities receiving the most green ARRA, whereas projects to improve energy efficiency or promote renewable energy are distributed more widely (online appendix table A2). Online appendix figure A5 shows the correlation between green (y axis) and nongreen (x axis) ARRA expenditure per capita for commuting zones with at least 25,000 inhabitants. The bivariate correlation between the two components of ARRA is positive and somewhat strong (0.393). As such, controlling for nongreen stimulus spending is important to accurately estimate the impact of green stimulus spending. We discuss our technique for doing so in section IV.

IV. Empirical Strategy

Our empirical strategy addresses two challenges unique to identifying the causal effect of the green stimulus. First, the green stimulus is small relative to the nongreen stimulus. Therefore, controlling for nongreen ARRA expenditures is essential since the recovery plan targeted markets hardest hit by the Great Recession. Second, the allocation of green investments may depend on structural characteristics of the local economy. We include several control variables designed to mitigate these threats to identification. Some controls describe each commuting zone's potential exposure and resilience to the Great Recession. Others capture the stringency of environmental policies in the local labor market as well as the relative importance of green versus nongreen employment in the local economy. Online appendix A2 describes these variables in more detail. However, areas receiving more green ARRA experienced higher employment growth before the Great Recession, even conditioning on these intervening factors. We address both of these issues in this section; IV.A introduces the main endogeneity issues to estimate the effect of green ARRA on employment, and IV.B discusses our approach to tackle them.

IV.A. Illustrating Endogeneity Issues

ARRA spending has been primarily designed to mitigate the effects of the Great Recession on local labor markets. Thus, it targets areas hardest hit

by the recession and is endogenous by construction. Controlling for nongreen ARRA expenditures is essential but potentially introduces another endogenous variable complicating the identification of the green ARRA effect (Angrist and Pischke 2008). The trade-off is between an error of misspecification from not including nongreen ARRA and a bias in estimating the green ARRA effect for including a bad control (nongreen ARRA) correlated with the error term. We address this by using a series of dummy variables for nongreen ARRA spending, which allows us to compare the effects of green ARRA in communities that received similar levels of nongreen ARRA investments.

To illustrate the difference in the allocation of green and nongreen ARRA as well as the source of data variation used for identification, we examine the distribution of the two types of spending along the nongreen ARRA distribution. Figure A6 in online appendix A1 reports the deviations from the mean and the standard deviation of green and nongreen ARRA spending per capita relative to the national mean for each vigintile of nongreen ARRA spending per capita. Since nongreen ARRA has been directed to areas hardest hit by the recession, the figure illustrates the extent to which green ARRA has been allocated following a different criterion. The left panel of figure A6 shows that the positive correlation between green and nongreen ARRA masks substantial variation across vigintiles as we observe commuting zones with low nongreen ARRA and high green ARRA or vice versa. In addition, the right panel suggests that the standard deviation of green ARRA within each vigintile is very similar across vigintiles with the exception of the first and last vigintile of nongreen ARRA spending. In our econometric analysis, we will use twenty dummies for nongreen ARRA vigintile to make sure that the effect of green ARRA is not capturing that of other ARRA programs. This particular functional form to treat nongreen ARRA allows testing the robustness of our results to the exclusion of vigintiles in which the dispersion of green ARRA spending is very high or low or the correlation with nongreen ARRA very high.

For green ARRA, identification is complicated by the presence of an additional source of endogeneity. Given the significant share of green ARRA spending devoted to long-term investments and research, the allocation of such spending may have followed criteria related to other structural features of the local economy such as the presence of a federal R&D laboratory or high-tech manufacturing. Thus, we directly explore the observable characteristics of a commuting zone that are correlated with green ARRA spending. Strong unbalances in the observable characteristics of commuting zones receiving different amounts of green ARRA are a red

spy of an unbalanced distribution also in unobservables (Altonji, Elder, and Taber 2005). We consider the association between the log of green ARRA spending per capita and two sets of covariates that will be used also as controls in our econometric model presented in the next section. The first set captures the economic conditions in commuting zone *i* before the Great Recession and are quite standard in the literature evaluating the ARRA (Wilson 2012; Chodorow-Reich and others 2012; Chodorow-Reich 2019).[13] The second set of variables are more specific to the green economy, such as the stringency of environmental regulation in the local area (Greenstone 2002), wind and solar energy potential (Aldy 2013), and an index of local green general skills (Vona and others 2018).[14] We also consider two alternatives to model regional fixed effects: year-varying census division dummies and state dummies. The choice of the way of modeling time-varying regional effects is nontrivial. Using census division dummies is a popular choice of previous ARRA evaluations, thus it allows us to compare our findings with previous literature (Dupor and Mehkari 2016). Using state fixed effects better accounts for state-specific policies important

13. We consider both the level and the pre-trends (2005–2007) in several variables such as total employment, unemployment, and employment in different sectors. As in Wilson (2012), we include the pre-sample level (average 2006–2008) and long pre-trends (2000–2007) for the following variables: total employment, employment in health and education, employment in the public sector, employment in manufacturing, construction, and extraction, unemployment. We also add other confounders of local labor market conditions such as pre-sample income per capita, a dummy equal one for commuting zones with positive shale gas production and import penetration. See online appendix A2 for details on data sources and construction of these variables.

14. As in Greenstone (2002), we use changes in the attainment status to National Ambient Air Quality Standards (NAAQS) for the six criteria air pollutants defined by the US Clean Air Act (CAA). We classify as nonattainment commuting zones in which at least one-third of the population resides in nonattainment counties. We also add a dummy variable to identify areas with nonattainment status for at least one of the NAAQS in 2006 and that therefore were already exposed to stringent CAA regulation. Since wind and solar energy received other types of support from the federal and state governments, including tax credits and loan guarantees as part of ARRA (Aldy 2013), we add proxies for the wind and solar potential interacted by year fixed effects. We include a dummy equal one for areas hosting a public R&D lab and the log of local population, as Vona, Marin, and Consoli (2019) show that is highly correlated with the size of the green economy in metropolitan areas. Since some green ARRA subsidies were allocated to state governments to be used throughout the state, we include a dummy variable indicating if the commuting zone includes a state capital. Finally, to proxy for the green capabilities of each commuting zone, we include the share of employment before the Great Recession (e.g., in 2006) in each commuting zone in occupations above the 75th percentile of the national distribution of GGS requirements, that is, skills most relevant in green jobs; see Vona and others (2018) and online appendix A2 for details on the green skill measures. See online appendix A2 for details on data sources and construction of these variables.

to a green economy, such as renewable portfolio standards, as well as for unobserved shocks that are geographically concentrated.

Online appendix table B1 shows that the inclusion of the vigintiles of nongreen ARRA is not enough to eliminate differences in observable characteristics that are significantly correlated with the intensity of green ARRA spending per capita. The table also highlights the different potential sources of endogeneity in the allocation of green ARRA: commuting zones receiving more green subsidies are both stronger in terms of technological expertise (workforce skills for the green economy, higher share of manufacturing employment, and the presence of a federal R&D lab) and somewhat more vulnerable to the Great Recession (i.e., higher share of employment in construction, which was particularly badly hit by the Great Recession). Areas receiving more green ARRA also have a larger share of employment in the public sector. However, in section V we confirm that our results are not driven by public sector employment.

The last diagnostic concerns violations of the parallel trend assumption. While untestable, the presence of pre-trends is typically used as a reliable diagnostic to assess the plausibility of this assumption. In our context, we test the possibility that employment growth before the Great Recession differs depending on the level of green ARRA received using an event study framework by including observations from 2000 to 2007. Since per capita green ARRA is correlated with the structural characteristics of the areas that are usually associated with sustained employment growth, we include controls for the observable commuting zone characteristics and vigintiles of nongreen ARRA described above. Observing that green ARRA went disproportionately to areas growing faster before the Great Recession conditional on these covariates indicates the presence of unobserved variables that are correlated with both the allocation of green ARRA and employment dynamics.

In sum, while the role of unbalances in the covariates can be mitigated by directly testing the robustness of the results to the exclusion of specific areas (i.e., those with R&D labs), the presence of pre-trends in some cases requires greater care to provide an accurate estimate of the effect of green ARRA on employment. We discuss the possible solution to this problem in the next section.

IV.B. Estimating Equation and Addressing Pre-trends

Our main econometric model is an event study model that jointly estimates the effects of green ARRA for years before and after the crisis. It can be seen as a straightforward extension of the econometric model

used in the papers reviewed by Chodorow-Reich (2019) that primarily exploits the cross-sectional variation of fiscal spending to identify local multiplier effects. The first main advantage of this approach is that we can explicitly tackle the potential pre-trends discussed above. The second advantage is being able to assess whether the effect of green ARRA lasts beyond the stimulus period, possibly generating a virtuous circle of (green) investments. Our dependent variable is the long difference between our measures of per capita employment in year t relative to our base year of 2008.[15] We estimate the following equation for a balanced panel of 587 commuting zones:

$$(1)\qquad \Delta \ln\left(y_{it}\right) = \alpha + \sum\nolimits_{\tau} \beta_{\tau} \ln\left(\frac{GreenARRA_i}{pop_{i,2008}}\right) \times D_{\{t=\tau\}} + \sum\nolimits_{\tau} X'_{it_0} \varphi_t \times D_{\{t=\tau\}} + \sum\nolimits_{\tau} G'_{it_0} \vartheta_t \times D_{\{t=\tau\}} + \mu_{i\in v,t} + \eta_{i\in c,t} + \epsilon_{it},$$
$$\tau = T_0, \ldots, 2006, \ldots, 2017$$

where $\epsilon_{i,t}$ is an error term, G'_{it_0} is a matrix of control variables specific to the green economy, and X'_{it_0} is a matrix of control variables used in previous ARRA evaluations (see footnotes 11 and 12 for details); $\mu_{i\in v,t}$ are year-specific dummy variables for the vigintiles of nongreen ARRA spending and $\eta_{i\in c,t}$ are year-specific region fixed effects (FE) at either the census division or state level. The dummy variables $D_{\{t=\tau\}}$ allow us to estimate separate coefficients for each year. The initial year T_0 is 2000 for total employment and 2005 for employment in different occupations and sectors.[16]

The main variable of interest is green ARRA spending, also rescaled by total population in 2008 in the commuting zone. While effective green spending spanned several years between 2009 and 2012, nearly all outlays were announced in 2009 (Wilson 2012). Because it is difficult to disentangle the announcement effect from the real spending effect, we build a time invariant measure of green spending as the total spending across those four years. We take a log transformation for both our dependent and main explanatory variable to account for the skewness in their respective

15. Employment is either green employment, total employment, or employment in a particular sector (construction, manufacturing, etc.) or occupation (managers, manual workers, etc.). See online appendix A2 for more details on data sources and measurement of our dependent variables.

16. American Community Survey micro data from the Integrated Public Use Microdata Series for the years 2001 to 2004 are only geo-localized at the state level, preventing us from building CZ-level indicators of different types of employment in these years.

distributions.[17] Therefore, our estimated coefficient $\hat{\beta}_\tau$ provides estimates of the number of jobs created by green ARRA in a given year.[18]

In all regressions, we cluster standard errors at the state level, using the state of the main county in each commuting zone. We cluster at the state level because the boundaries of a local labor market can be larger than the commuting zone perimeter, especially in post-recession times when workers are forced to search for a job in a larger area. Finally, we weight observations using the population level in 2008.

For ease of interpretation, we also present an alternative specification where we allow the coefficient of green ARRA and of all the other covariates, including region fixed effects and the vigintiles for nongreen ARRA, to vary only among four symmetrical periods: early pre-ARRA (2000–2003), late pre-ARRA (2004–2007), the short run (2009–2012), and the long run (2013–2016).[19] For samples beginning in 2005, only one preperiod is included. We estimate:

$$(2)\qquad \Delta \ln\left(y_{it}\right) = \alpha + \sum\nolimits_\tau \beta_\tau \ln\left(\frac{GreenARRA_i}{pop_{i,2008}}\right) \times D_{\{t=\tau\}} + \sum\nolimits_\tau X'_{it_0}\varphi_t \times D_{\{t=\tau\}} + \sum\nolimits_\tau G'_{it_0}\vartheta_t \times D_{\{t=\tau\}} + \mu_{i\in v,t} + \eta_{i\in c,t} + \epsilon_{it},$$

where τ = early pre-ARRA, late pre-ARRA, short run, long run.

Here $\hat{\beta}_\tau$ indicates the average number of job-years created by green ARRA in each of these four periods. So that the value can always be interpreted as growth in employment relative to 2008, in these models we define the dependent variable as follows:

$$\Delta \ln\left(y_{i,t}\right) = \ln\left(\frac{y_{i,2008}}{pop_{i,2008}}\right) - \ln\left(\frac{y_{i,t}}{pop_{i,2008}}\right) = \ln\left(\frac{y_{i,2008}}{y_{i,t}}\right) \qquad \text{if } t < 2008$$

$$\Delta \ln\left(y_{i,t}\right) = \ln\left(\frac{y_{i,t}}{pop_{i,2008}}\right) - \ln\left(\frac{y_{i,2008}}{pop_{i,2008}}\right) = \ln\left(\frac{y_{i,t}}{y_{i,2008}}\right) \qquad \text{if } t > 2008$$

17. As show in online appendix table B2, using logs reduces sensitivity to outliers due to the skewed distribution of green ARRA. In the robustness checks, we show that the log-log results do not change when removing outliers. In contrast, if levels of all variables are used, we only estimate a positive effect of green ARRA on employment if these outliers are dropped from the sample.

18. Because of the log specification, the quantification of the number of jobs created is not straightforward as in related papers. In online appendix C we report the arithmetic to translate the estimated coefficients into number of jobs created.

19. In this specification, we drop 2017 to make the length of each time period equal.

A further advantage of this specification is that it reduces the number of coefficients to be estimated, which is important for assessing the role of mediating factors of green ARRA effects. In section V, we consider whether the existing skill composition in each commuting zone changes the effectiveness of green ARRA, focusing on the mediating effect of a preexisting pool of workers with a high level of green skills. In doing so, we add to equation (2) a full set of interactions between the green skill index and the ARRA spending in the four periods.

Constraining the coefficients to remain constant within each subperiod also provides us one strategy to address possible pre-trends. Recall that, given the unbalances in the covariates discussed in the previous section and the possible presence of pre-trends discussed earlier, we cannot assume that the allocation of green ARRA spending to commuting zones is quasi random, even after including our rich set of controls. The pre-trend effect $\hat{\beta}_{pre}$ reflects the presence of unobserved variables that are correlated with both the allocation of green ARRA and the outcome variables. When pre-trends are present (i.e., $\hat{\beta}_{pre} > 0$), we compute the long- and short-term effect of green ARRA by subtracting its effect before 2008 in a period that is equally distant to the benchmark year as, by construction, effects will get closer to zero as we approach the base year of 2008 when pre-trends are present. Thus, $\hat{\beta}_{short} - \hat{\beta}_{pre2004–07}$ and $\hat{\beta}_{long} - \hat{\beta}_{pre2000–03}$ can be interpreted as the net effect of green ARRA on total jobs created per year in the short or long run, respectively. As will be extensively discussed in section V, the credibility of these net effect estimates rests upon an untestable assumption regarding the functional form of the relationship between employment and green ARRA before the Great Recession.

IV.C. An Alternative Instrumental Variable Strategy

As an alternative identification strategy, we exploit the well-known fact that ARRA spending was allocated according to formulas that were in use before the passage of the act (Chodorow-Reich 2019).[20] Importantly, the

20. According to Conley and Dupor (2013), two-thirds of ARRA spending was allocated using such a formulaic approach to privilege shovel-ready projects that have an immediate impact on the economy. For instance, spending in road construction, education, and health were allocated by the act using the formulas in place before the act (Wilson 2012; Garin 2019). An example for green ARRA are Energy Efficiency and Conservation Block Grants. This program was created by the Energy Independence and Security Act of 2007, which provided specific guidelines for distribution of funds. ARRA provided additional funding for this program and stipulated that the same formulas for eligibility in the 2007 act be used (American Recovery and Reinvestment Act of 2009). However, many DOE ARRA projects supported new infrastructure, such as grid modernization, and do not appear to have been allocated formulaically.

formulaic instrument has a typical shift-share structure used in the seminal literature on cross-sectional multipliers (Nakamura and Steinsson 2014). In previous studies, such an instrument satisfied the exclusion restriction of affecting total employment only through ARRA spending because the main source of endogeneity was the local effect of the Great Recession.

Following these studies, we use an instrument that combines the initial "share" of EPA plus DOE spending in the commuting zone (over total DOE and EPA spending) with the green ARRA "shift." Such an instrument adds an exogenous shock in green expenditures to areas that were already receiving a larger amount of green spending before ARRA.[21] Unfortunately, endogeneity of green ARRA is also related to the persistent effect of pre-ARRA green investments by both private and public institutions. Thus, this instrumental variable strategy is less effective in our case. Because such an instrument adds an exogenous shock in green expenditures to areas that were already receiving larger green investments before ARRA, we face a problem similar to that put forward by Jaeger, Ruist, and Stuhler (2018), who note that a shift-share instrument conflates short- and long-term effects. We follow their suggestion and take a share far in the past (i.e., an average share of DOE plus EPA spending between 2003 and 2004), under the assumption that the effect of past spending gradually fades away and thus is excludable from the second stage. Note that having a reliable measure of pre-ARRA green government spending would be the ideal solution to distinguish the additional contribution of green ARRA from that of past trends associated with pre-ARRA green spending. However, as explained in online appendix D, building an accurate measure of pre-ARRA green spending is difficult due to the lack of details in public spending data pre-ARRA. Summing up, while the instrumental variable (IV) mitigates endogeneity related to nonrandom assignment of green ARRA subsidies, it captures the effect of past and present green ARRA on areas that were already on a green path and thus it is not well suited to mitigate violations of the nonparallel trend assumption mentioned above.

21. The instrument of green ARRA reads as $IV_i = \frac{DoE\ Pre{-}ARRA_{i,2003-04}}{DoE\ Pre{-}ARRA_{2003-04}} \times \frac{Green\ ARRA\ DoE}{Pop_{2008}} + \frac{EPA\ Pre{-}ARRA_{i,2003-04}}{EPA\ Pre{-}ARRA_{2003-04}} \times \frac{Green\ ARRA\ EPA}{Pop_{2008}}$, where total green ARRA EPA and DOE per capita is reallocated to commuting zones depending on their respective pre-ARRA shares of spending over the national total, that is, $\frac{DoE\ Pre{-}ARRA_{i,2003-04}}{DoE\ Pre{-}ARRA_{2003-04}}$ and $\frac{EPA\ Pre{-}ARRA_{i,2003-04}}{EPA\ Pre{-}ARRA_{2003-04}}$. See online appendix D for further details.

We directly test this conjecture by running the two-stage least squared counterpart of equation (2) using the instrument described above. The detailed results are illustrated in online appendix D for sake of space. It is worth noting first that the predictive power of the shift-share instrument is poor, with an *F*-test of the excluded instrument just above the usual cutoff of 10 (for census dummies) and just below that cutoff (for state dummies; see table D1). The weak instrument problem is consistent with the fact that DOE spending (the bulk of green spending) was redirected toward green programs. Second, compared to the ordinary least squares (OLS) estimator, the IV largely overstates both the pre-trends for total employment ($\hat{\beta}_{pre}$; see table D2) and the long-term effect of green ARRA per capita ($\hat{\beta}_{long}$). This latter result is consistent with previous studies on fiscal multipliers that found a larger job creation effect when a credible instrument is used (Nakamura and Steinsson 2014). Although the IV results suggest that the effect of green ARRA is highly heterogeneous and much stronger on compliers, they also exacerbate the source of endogeneity associated with the presence of pre-trends.[22] The OLS estimate is probably more conservative (being the average of the exogenous shock on green compliers and the endogenous shock on noncompliers) but also more reliable as less likely to conflate the effect of green ARRA with that of past green policies. Therefore, in the remainder of this paper, we choose OLS as the preferred estimator.

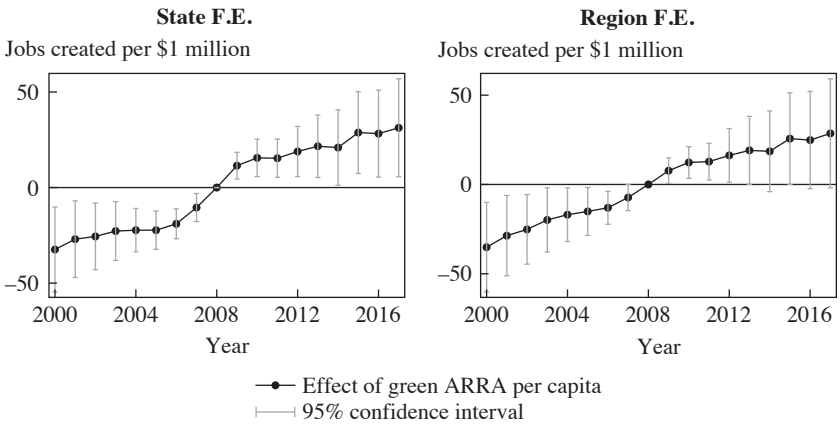
V. Results

This section presents the main results of the paper. We begin with total employment, including a discussion of the time profile of employment trends and the mediating effects of green skills. We then explore green ARRA restructured labor markets by looking at different occupations and sectors.

Figure 2 plots the year-by-year effects of green ARRA using both state and regional fixed effects (FE). The time profiles of both are similar, except that the pre-trends are nearly constant during the early pre-ARRA period (2000–2003) using state fixed effects, which also provide more precise estimates of the yearly job creation. Because state fixed effects also more effectively control for other green policies and unobserved characteristics correlated with employment growth, we consider the state fixed effect

22. Indeed, the IV may capture the effect of past and present green ARRA on areas that were already on a green path, that is, compliers in a local average treatment effect (LATE) interpretation (Imbens and Angrist 1994).

Figure 2. Total Employment: Jobs per Year Created by $1 Million Green ARRA



Source: Authors' calculations.

specification as our preferred one. In both cases, we see early job creation that continues through our sample period, leveling off at the end. By the end of the sample, an additional $1 million green ARRA creates just over thirty-two jobs per year.

While these results are suggestive of permanent job creation, they must be interpreted carefully because of the violation of the parallel trend assumption. We turn to the model estimating average effects for each of four four-year periods to illustrate how the presence of preexisting employment trends affects the interpretation of our results. The estimates ($\hat{\beta}_{short}$ and $\hat{\beta}_{long}$) provide the *gross* average short-run and long-run effects. Using the specification of equation (2), we can estimate the net average jobs per year created by subtracting a pre-trend effect constrained to be constant within two subperiods (i.e., 2000–2003 and 2004–2007). Before discussing these results in table 2, we use figure 2 to gain insights on the implicit assumption that pre-trend effects change mostly between (rather than within) subperiods. In figure 2, the slope connecting two neighboring years represents the annual growth of per capita employment for a community receiving an extra $1 million green ARRA. Consistent with the fact that state fixed effects do a better job controlling for unobserved characteristics correlated with pre-trends, this slope is flat until 2005, and it is primarily between 2006 and 2008 that communities receiving more green

Table 2. Estimates for Total Employment

Change in log employment per capita compared to 2008	*State FE*	*Census division FE*
Green ARRA per capita (log) × D2000–2003	0.0051***	0.0052**
	(0.0017)	(0.0020)
Green ARRA per capita (log) × D2004–2007	0.0034***	0.0025**
	(0.0007)	(0.0009)
Green ARRA per capita (log) × D2009–2012	0.0029***	0.0023**
	(0.0008)	(0.0009)
Green ARRA per capita (log) × D2013–2016	0.0045***	0.0039*
	(0.0017)	(0.0020)
Jobs per year created, $1 million green ARRA		
Pre-ARRA (2000–2003)	27.15***	27.65**
	(8.88)	(10.47)
Pre-ARRA (2004–2007)	18.88***	13.52**
	(3.95)	(5.21)
Short run (2009–2012)	15.3***	12.32**
	(4.31)	(4.62)
Long run (2013–2016)	25.52***	22.26*
	(9.41)	(11.49)
Short (2009–2012) to Pre (2004–2007)	–3.03	–0.82
	(3.72)	(4.50)
Long (2009–2012) to Pre (2000–2003)	–3.31	–7.1
	(7.43)	(9.84)
R^2	0.7688	0.7032
Observations	9979	9979

Source: Authors' calculations.

Note: Regressions weighted by CZ population in 2008. Sample: 587 commuting zones with at least 25,000 residents in 2008. Year fixed effects and state (or census division) times period fixed effects included. Additional control variables (interacted with D2000–2003, D2004–2007, D2009–2012, and D2013–2016 dummies): vigintiles of nongreen ARRA per capita, share of employment in occupations with GGS > p75 (2006), population 2008 (log), income per capita (2005), import penetration (year 2005), pre-trend (2000–2007) employment manufacturing / population, pre-trend (2000–2007) total employment / population, pre-trend (2000–2007) employment construction / population, pre-trend (2000–2007) employment extractive / population, pre-trend (2000–2007) employment public sector / population, pre-trend (2000–2007) unemployed / population, pre-trend (2000–2007) employment education and health / population, total employment (average 2006–2008) / population, employment manufacturing (average 2006–2008) / population, employment construction (average 2006–2008) / population, employment extractive (average 2006–2008) / population, employment public sector (average 2006–2008) / population, unemployed (average 2006–2008) / population, employment education and health (average 2006–2008) / population, shale gas extraction in commuting zone interacted with year dummies, potential for wind energy interacted with year dummies, potential for photovoltaic energy interacted with year dummies, federal R&D lab, commuting zone hosts the state capital, nonattainment CAA old standards, nonattainment CAA new standards. Data on total employment and employment share by industry come from BLS-QCEW. See online appendix A2 for details. Standard errors clustered by state in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

ARRA experienced faster employment growth.[23] Since green spending was allocated to areas more resilient to the Great Recession, we can say very little on short-term effects either subtracting the pre-trends or not. The absence of a net short-term effect can either reflect a fast convergence to a higher precrisis steady state (so it should be interpreted as evidence supporting the use of green spending to restart the economy) or the greater resilience of greener areas (so it should be interpreted as evidence of lack of additionality).

In turn, our analysis allows us to set credible bounds to the long-term effect. The flat job-year effect in the early precrisis period indicates that $\hat{\beta}_{pre2000–03}$ reflects the steady-state level of the employment gap in favor of communities receiving more green ARRA that results from greater resilience to the recession afterward. This assumes that this greater resilience continued through the post-ARRA period. Consequently, the difference between $\hat{\beta}_{long}$ and $\hat{\beta}_{pre2000–03}$ is the lower bound of the employment effect.[24] For the upper bound, we use the gross job creation, $\hat{\beta}_{long}$. This upper bound is plausible for at least two reasons. First, previous literature shows that the OLS estimates of local job multipliers are smaller than the IV estimates. This finding is further corroborated when we let the impact of green ARRA vary depending on the endowment of green skills in a commuting zone. Second, with an accommodating monetary policy and spillovers across regions, a local spending effect typically underestimates the aggregated effect (Chodorow-Reich 2020; Dupor and McCrory 2018).

Table 2 presents these results. Both the gross average short-run and long-run effects ($\hat{\beta}_{short}$ and $\hat{\beta}_{long}$) are positive and statistically different from zero. In terms of gross job creation, $1 million of green spending adds between 12.3 (with census FE) and 15.3 (with state FE) new jobs per year in the short term and 22.3 (with census FE) and 25.5 (with state FE) jobs per year in the long term. Using census division fixed effects for comparability, the short-term estimates are in the lower range of estimates of

23. These results are confirmed by using the first difference of employment as the dependent variable, allowing us to explicitly test for changes in the annual growth rate in employment. We also check the resilience of different commuting zones using monthly data near the peak of the Great Recession. Results confirm that greener areas were more resilient to the crisis. Both results are available upon request from the authors.

24. Note that $\hat{\beta}_{pre2000–03}$ is greater than $\hat{\beta}_{pre2004–07}$ even if pre-trend accelerates in the latter period. The reason is that the $\hat{\beta}_{\tau}$ is forced to be zero in 2008. Thus, subtracting $\hat{\beta}_{pre2000–03}$ allows us to retrieve a true lower bound of the net effect.

15985-01a-Popp-4thPgs.indd 26 5/12/22 12:11 PM

papers evaluating other programs of ARRA (Chodorow-Reich 2019).[25] For the long-term estimates, there is no clear benchmark in the literature using the cross-sectional approach to estimate local multipliers. An exception is Garin (2019), who finds no long-term effect of construction spending under ARRA. Green spending may have a larger and more persistent long-term effect on job creation than other types of infrastructural investment.

However, this interpretation is complicated by the fact that the net average effect (our lower bound) is statistically indistinguishable from zero. Unlike in Garin (2019), where pre-trends were not an issue, the presence of pre-trends produces a plausible estimation range between zero and a statistically significant estimate of 25.5 jobs per year. The large range of estimates for the long-term green spending multiplier does not allow us to confidently support the statement that green spending is a driver of permanent jobs. At a minimum, the lower bound indicates that green spending helped areas that were relatively better off before the Great Recession maintain their precrisis advantage. To lend support to the interpretation that green ARRA picked winners, we examine next the mediating effect of green skills.

V.A. The Mediating Effect of Green Skills

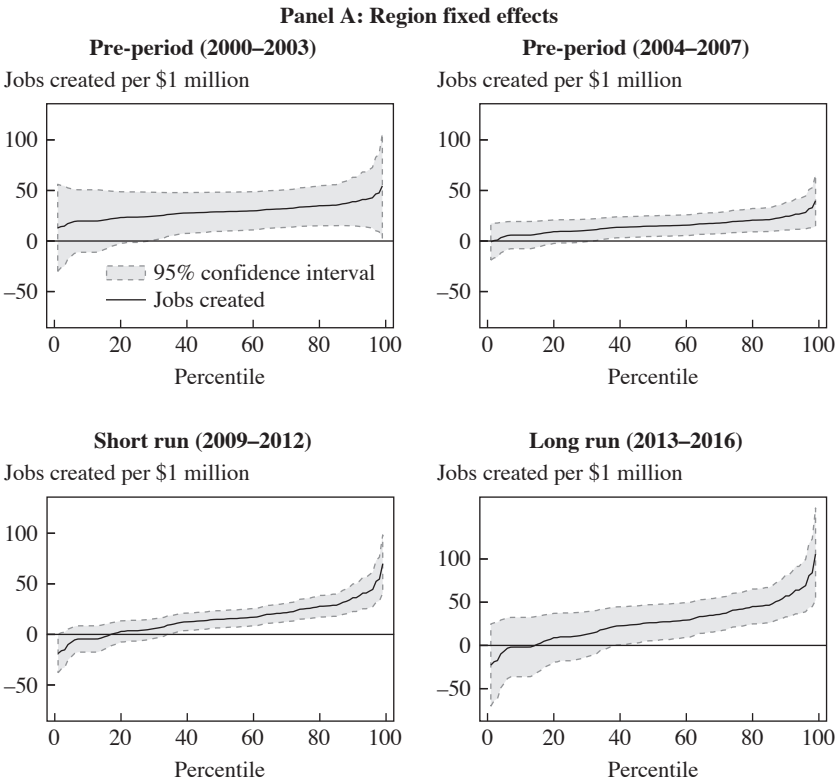
The large uncertainty on the aggregate effect may mask substantial heterogeneity across communities with different levels of green competences. Commuting zones with a workforce more prepared to perform green tasks are more likely to experience larger gains in both the short and long run. Looking at the heterogeneous effect with respect to the existing skill base of the workforce allows us also to shed light on the large gap between the OLS and IV estimates, improving the interpretation of our results. Because the IV results highlight much larger effects on compliers, that is, commuting zones already investing in the green economy, this result reinforces our expectation that the green stimulus would be more effective in areas with a higher concentration of green skills. Recall, indeed, that the initial concentration of green skills in a region is positively associated with the allocation of green ARRA spending.

25. Note that other papers estimate gross job creation effects, while we privilege the hyperconservative estimation given by the net short-term effect. As discussed in section IV, other papers also use a formulaic IV that identifies the effect on compliers, which is found to be generally larger than the effect on the entire population.

In table 1, we showed that the types of skills workers need to work in green jobs are different than the skills needed in the rest of the economy, requiring more on-the-job training as well as engineering and technical competences. We use the data on green skills described in section II to identify the share of employment in each commuting zone in occupations with green general skill (GGS) importance in the 75th percentile or higher in 2006 (i.e., prior to the recession). This includes 113 occupations, listed in table A7 in online appendix A2. While these jobs need not themselves be green, this captures the local endowment of the types of skills in high demand in a green economy.

We augment the model constraining the effects across four periods by interacting our green ARRA variables (early-pre, late-pre, short, and long) with the share of employment in occupations with GGS importance in the 75th percentile or higher. Note that all models already control for the initial concentration of green skills in a region and allow the effect of the concentration of green skills to vary across periods. Figure 3 shows the marginal effect of green ARRA at different levels of initial green skills for the specification with both state and census division dummies. Complete regression results are given in table B3 of online appendix B. The results show, as expected, the importance of the initial skill base. Gross job creation increases with the level of green skills. Comparing communities at the median and 75th percentiles of green skills, 40 to 70 percent more jobs are created in communities with high green skills in both the short and long run. Using the preferred specification with state fixed effects, short-run gross job creation becomes significant when the share of workers with high green skills reaches the 20th percentile of all communities, or 23.4 percent (17th percentile, or 22.3 percent, with regional fixed effect). Long-run gross job creation becomes significant when the share of workers with high green skills reaches the 32nd percentile of all communities, or 24.0 percent (39th percentile, or 24.9 percent, with regional fixed effects). Importantly, while we still observe pre-trends, the relationship between pre-trends and green skills appears weaker, particularly in the model using regional fixed effects. The magnitude of the interaction coefficient between green skills and green ARRA is between two and three times larger in the post-ARRA period than in the pre-ARRA period (online appendix table B3). Thus, green spending enhances net job creation as well. Because of the added noise when subtracting pre-trends, estimates of net job creation are imprecise. Short-run net job creation becomes positive at the 73rd percentile of all communities, but never statistically significant (46th percentile with regional fixed effects, but not statistically significant until reaching the

Figure 3. Variation in the Effect of Green ARRA on Employment by Initial Green Skills



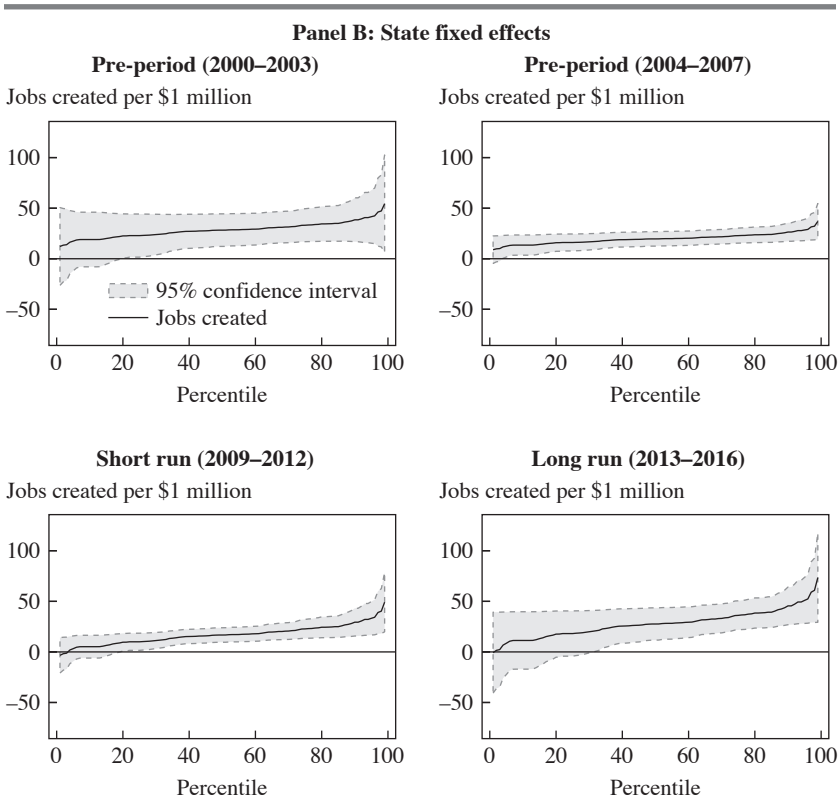
(*continued on next page*)

97th percentile). In the long run, the net effect of jobs created becomes positive in the 72nd percentile (64th with regional fixed effects) but is never statistically significant. Interestingly, by allowing an easier match between green tasks and skills, the availability of green skills increases the short-run effectiveness of green spending more than the long-run effectiveness.[26]

These results are even more remarkable when noting that the initial share of occupations in the upper quartile of GGS importance itself is a

26. To verify that this effect is not simply driven by the fact that communities with more green skills are richer, in online appendix table B4 we conduct a falsification test by interacting green ARRA with per capita income in 2005, rather than green skills. The interaction is never significant, and the estimates of jobs created hardly vary across different income levels.

Figure 3. Variation in the Effect of Green ARRA on Employment by Initial Green Skills (*Continued*)



Source: Authors' calculations based on estimates from online appendix table B3.

good predictor of future employment growth, with an effect that increased after the crisis (online appendix table B3). Indeed, in the most conservative specification with state fixed effects, a one standard deviation in the green skills share (0.027) accounts for a 1.1 percent difference in employment growth in the early precrisis period that increases up to 2.0 percent in the short term and 2.3 percent in the long run (although with a *p*-value of 0.101).[27] Online appendix table B1 shows that the initial share of

27. With regional fixed effects, the acceleration is much more pronounced: a standard deviation in the share of green skills explains only 1.4 percent difference in employment growth in the early precrisis period, while it accounts for a differential employment growth of 3.3 percent in the short run and 4.1 percent in the long run (p-value = 0.013).

15985-01a-Popp-4thPgs.indd 30 5/12/22 12:11 PM

occupations in the upper quartile of GGS importance is also strongly correlated with the allocation of green ARRA subsidies. In combination, these results reinforce our interpretation of the green stimulus as a successful example of picking the winners. The main policy lesson is that increasing the green skills in a community, such as through job training focused on mid-level technical and engineering skills, should represent a key part of a successful policy package for the green transition.

V.B. Reshaping the Economy: Heterogeneous Effects of Green ARRA

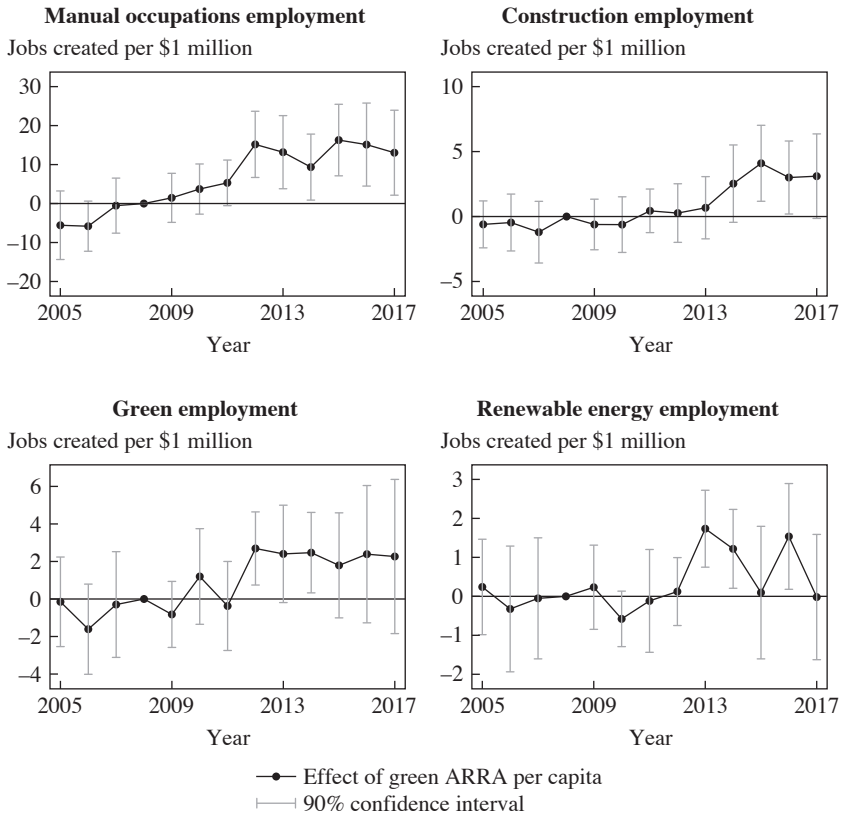
To summarize our results on total employment, we find evidence of gross job creation that is strongest in communities with more preexisting green skills. However, the presence of pre-trends in the total employment regressions complicates the interpretation of these results. That is not the case when looking at specific types of workers, where, as will be shown, we find no evidence of pre-trends complicating the interpretation. Here, we show that green ARRA plays an important role in *reshaping* the economy. We present results for all manual labor, construction, overall green employment, and renewable energy employment. We concentrate on results using the preferred specification with state fixed effects.

The choice of these four categories of jobs is consistent with the objectives of the green ARRA stimulus. Clearly, the creation of green and renewable energy jobs is the main channel through which the effect of green ARRA spending should take place (Vona, Marin, and Consoli 2019).[28] As noted by Garin (2019), a necessary condition for a positive effect of specific government spending is that it should create jobs in the sectors most likely affected by such spending. We then focus on manual labor and construction employment, both because green spending includes investments in infrastructure and construction (such as improving building energy efficiency) and for the importance of these sectors in the debate on the distributional effects of trade and technology shocks (Autor, Dorn, and Hanson 2013; Acemoglu and Restrepo 2020) and of the rise of populism in the United States (Autor and others 2020).

Figure 4 shows year-by-year results for specific employment types, while table 3 complements these results by reporting the point estimates of the green ARRA coefficients for the pre-ARRA period ($\hat{\beta}_{pre}$), the short

28. Green employment is measured by reweighing occupational employment by the share of specific tasks in each occupation that O*NET defines as green; see online appendix A2 and Vona, Marin, and Consoli (2019).

Figure 4. Jobs per Year Created by $1 Million Green ARRA



Source: Authors' calculations.
Note: All models estimated using state fixed effects.

term ($\hat{\beta}_{short}$), and the long term ($\hat{\beta}_{long}$). The table also reports the number of jobs per year created per millions of dollars spent. Since pre-trends are not an issue here, we only include jobs created for the gross effects ($\hat{\beta}_{short}$ and $\hat{\beta}_{long}$).[29]

For manual labor, figure 4 reveals the presence of a slightly upward pre-trend that is, however, never statistically significant.[30] Statistically significant

29. Subtracting insignificant pre-trends simply adds noise to the interpretation without providing new information.

30. Moreover, the results for manual labor are robust to using state or region fixed effects, and there is clearly no evidence of a pre-trend using regional fixed effects (online appendix table B1).

15985-01a-Popp-4thPgs.indd 32 5/12/22 12:11 PM

Table 3. Reshaping Results (State Fixed Effects)

Change in log employment (by type) per capita compared to 2008	*Manual occupations*	*Construction employment*	*Green employment*	*Renewable employment*
Green ARRA per capita (log) × D2005–2007	0.0028	0.0023	0.0028	0.0003
	(0.0023)	(0.0034)	(0.0043)	(0.0083)
Green ARRA per capita (log) × D2009–2012	0.0051**	–0.0004	0.0027	–0.0013
	(0.0023)	(0.0043)	(0.0041)	(0.0061)
Green ARRA per capita (log) × D2013–2016	0.0102***	0.0100**	0.0086*	0.0164**
	(0.0036)	(0.0045)	(0.0051)	(0.0076)
Jobs per year created, $1 million green ARRA				
Pre-ARRA (2005–2007)	3.97	0.71	0.71	0.03
	(3.26)	(1.03)	(1.08)	(0.69)
Short-run (2009–2012)	6.17**	–0.1	0.65	–0.09
	(2.75)	(1.00)	(0.98)	(0.42)
Long-run (2013–2016)	13.4***	2.54**	2.33*	1.19**
	(4.75)	(1.15)	(1.37)	(0.56)
R^2	0.5849	0.7177	0.3864	0.2786
Observations	7044	7044	7044	7044

Source: Authors' calculations.

Note: Regressions weighted by CZ population in 2008. Sample: 587 commuting zones with at least 25,000 residents in 2008. Year fixed effects and state (or census division) times period fixed effects included. Additional control variables (interacted with D2000–2003, D2004–2007, D2009–2012, and D2013–2016 dummies): vigintiles of nongreen ARRA per capita, share of employment in occupations with GGS > p75 (2006), population 2008 (log), income per capita (2005), import penetration (year 2005), pre-trend (2000–2007) employment manufacturing / population, pre-trend (2000–2007) total employment / population, pre-trend (2000–2007) employment construction / population, pre-trend (2000–2007) employment extractive / population, pre-trend (2000–2007) employment public sector / population, pre-trend (2000–2007) unemployed / population, pre-trend (2000–2007) employment education and health / population, total employment (average 2006–2008) / population, employment manufacturing (average 2006–2008) / population, employment construction (average 2006–2008) / population, employment extractive (average 2006–2008) / population, employment public sector (average 2006–2008) / population, unemployed (average 2006–2008) / population, employment education and health (average 2006–2008) / population, shale gas extraction in commuting zone interacted with year dummies, potential for wind energy interacted with year dummies, potential for photovoltaic energy interacted with year dummies, federal R&D lab, commuting zone hosts the state capital, nonattainment CAA old standards, nonattainment CAA new standards. Employment data calculated by multiplying total employment (BLS-QCEW) by the share of workers in each category, taken from ACS. See online appendix A2 for details. Standard errors clustered by state in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

job creation for manual workers begins in 2012, with 14.9 jobs created. The level of jobs per year created remains steady after 2012, providing evidence of permanent job creation. On average, 13.4 manual labor jobs per year are created per $1 million green ARRA in the long run, compared to only 6.17 in the short run (table 3). Thus, over half of the gross total job creation was in manual labor.

Construction jobs exhibit no pre-trends and a similar time profile of the post-crisis effects, although job creation was delayed until 2015. This further lag in job creation is consistent with the fact that administrative delays tend to be longer for infrastructure spending (Ramey 2020). Compared to Garin's (2019) study of direct construction spending, our results both are larger and suggest permanent job creation, with an average of 2.54 construction jobs per year created in the long run (table 3). While this can be partly explained by fact that Garin's (2019) county-level analysis leaves out positive demand spillovers across counties, further research is required to explore the labor intensity of green infrastructure relative to other infrastructure. For instance, O*NET tasks data suggest that maintenance activities are likely to be particularly important for green jobs, suggesting a possible channel through which jobs are maintained in areas with energy-efficient buildings or renewable energy infrastructure.

For green employment, the results are estimated much less precisely. Still, the absence of pre-trend is very clear, suggesting that areas receiving more green ARRA were not necessarily experiencing a faster green growth before the Great Recession. Significant green job creation begins in 2012, with 2.8 jobs per year, and again persisted until the end of our sample. Note that, as the estimates of overall green employment are imprecise, the estimate of 2.3 jobs created in the long run is only significant at the 10 percent level (table 3). While this effect seems small, it is roughly 10 percent of gross total job creation, whereas green employment is just 4.7 percent of total employment from 2005 to 2017 in our data.[31] For renewable energy jobs, the patterns largely mimic those observed for green employment with two caveats. On the one hand, the year-by-year estimates plotted in figure 4 are very unstable and imprecise. On the other hand, the reshaping effect emerges more clearly as highlighted by the positive and significant long-run effect (table 3). Not surprisingly, the effect is also large relative to the share of renewable energy employment (1.3 percent) with $1 million green ARRA creating about 1.2 renewable jobs per year in the long run.

To summarize, the reshaping effect of the green fiscal stimulus emerges clearly from our results. Of the gross job creation for total employment, half was in manual labor, while job creation in construction and green occupations each make up about 10 percent.[32] These findings are qualitatively

31. The 4.7 percent estimate is higher than the estimate of 3.1 percent provided by Vona, Marin, and Consoli (2019) for 2014. This is due to an aggregation bias. See online appendix A2 for details.

32. Note that we cannot sum up these figures as these jobs largely overlap.

confirmed in comprehensive robustness checks (see online appendix B), where we exclude areas with unbalanced characteristics, define green ARRA in different ways, and group areas with similar nongreen ARRA spending differently. Moreover, for green jobs our definition of green ARRA is conservative. If we exclude R&D spending and loans and contracts, which are more likely to support large infrastructure projects (leaving only grants), we obtain statistically significant estimates of up to 3.8 green jobs created in the long run (online appendix tables B9 and B10).

Online appendix table B11 provides additional insight to the heterogeneous effects of green ARRA. We look for employment effects in three additional sectors: manufacturing (NAICS 31-33), public administration (NAICS 92), and support services including waste management (NAICS 56). Those sectors are either most likely to receive green subsidies (e.g., waste management) or to employ workers needed to administer and monitor ARRA programs (e.g., public administration). While we observe gross job gains in manufacturing, these gains are offset by pre-trends in the 2005–2007 period. Green ARRA spending reduces the share of employment in the public sector, at least in the short run. This result reassures us that the effect on total employment is not associated with a crowding out of private jobs.

V.C. Distributional Effects on Manual Workers

While the green stimulus reshaped labor markets by increasing the size of the local green economy, the distributional effect of the stimulus among workers is less clear. In online appendix table B12 and figure B2 we estimate separate models for different broad groups of workers following a standard grouping in the literature on task-biased technological change (Acemoglu and Autor 2011): abstract occupations, service workers, clerical occupations, and manual labor (see table A8 in online appendix A2). The important result here is that all job creation from green ARRA occurs in manual labor occupations, while both the net and the gross effects for other occupational groups are far from being significant at conventional levels. Moreover, the results for abstract and clerical workers suggest additional insights into the pre-trend. Both show evidence that areas receiving more green ARRA grew faster right before the Great Recession. However, neither occupation group experiences long-run employment gains. The short-run gains for abstract employment end in 2010. While not definitive, these suggest that any pre-trends that continued past the Great Recession were unlikely to extend more than a couple of years, making the lower bound estimate of total employment appear less plausible. Although the upper

bound gross job creation may probably overstate job creation, it seems likely that green ARRA created at least some new jobs in the long run.

Manual workers have been losing in terms of wages and employability for trade (Autor, Dorn, and Hanson 2013), automation (Acemoglu and Restrepo 2020), and, but to a lesser extent, the effect of climate policies (Marin and Vona 2019). It is thus important to provide an in-depth look at how the green stimulus affected manual labor. Table 4 considers the effect of green ARRA on manual labor wages (columns 1–3) and on educational attainment of manual workers. First, column 1 replaces changes in per capita employment as the dependent variable with the average hourly wage of manual workers. Despite increasing demand for manual labor, green ARRA investments did not increase the wages of manual workers.[33] In columns 2 and 3, we see that most of the increase in manual labor jobs occurred in jobs where workers earned less than the US median wage for all manual workers. These missing wage gains highlight the well-known deterioration of the bargaining power of manual workers facing a series of correlated negative shocks. While the manual labor jobs created in the short run were evenly distributed between workers having or not having more than a high school education, by the long run most gains were among workers with less than a high school education (columns 4–5). Overall, public spending in the green economy under ARRA was too small to substantially improve the bargaining power of manual workers. Thus, it remains an open question whether larger green stimuli can revert the long-run decline of the working conditions of manual labor.

VI. Policy Discussion

Our results can inform both the design of future green fiscal stimuli programs and address longer-term concerns about job losses in the transition to a green economy. Our first key finding is the potential for green investments to reshape the economy. Over half of gross job creation was in manual labor, and green jobs also experienced a notable increase. Importantly, these jobs appear permanent, so that green investments offer new opportunities for workers often left behind by changing environmental regulations and by other structural transformations in the labor markets, such as trade and automation. For example, as discussed in section II, the skills

33. This may be explained by the need to comply with prevailing wage laws. Since contractors were required to document that workers were paid prevailing wages, they had little incentive to pay more than the prevailing wage. We thank Joe Aldy for this insight.

15985-01a-Popp-4thPgs.indd 37

Table 4. Focus on Manual Occupations

Dep var: Change in log employment (by category) per capita compared to 2008 (except column 1)	*Average hourly wage of manual workers*	*Manual workers, hourly wage > US med. for manual workers*	*Manual workers, hourly wage < US med. for manual workers*	*Manual workers with education > high school degree*	*Manual workers with high school degree or less*
Green ARRA per capita (log)	0.0080	0.0049	–0.0005	–0.0016	0.0044
× D2005–2007	(0.0056)	(0.0036)	(0.0031)	(0.0044)	(0.0029)
Green ARRA per capita (log)	–0.0045	0.0032	0.0093**	0.0094*	0.0041
× D2009–2012	(0.0048)	(0.0031)	(0.0036)	(0.0049)	(0.0025)
Green ARRA per capita (log)	–0.0010	0.0084*	0.0132***	0.0109**	0.0098**
× D2013–2017	(0.0052)	(0.0044)	(0.0049)	(0.0053)	(0.0042)
Jobs per year created, $1 million green ARRA					
Pre-ARRA (2005–2007)	N/A	3.63	–0.36	–0.59	4.67
		(2.66)	(2.08)	(1.58)	(3.04)
Short-run (2009–2012)	N/A	2.08	5.36**	3.24*	3.54
		(1.99)	(2.06)	(1.72)	(2.17)
Long-run (2013–2017)	N/A	5.92*	8.08***	4.27**	9.06**
		(3.09)	(3.01)	(2.09)	(3.90)
R^2	0.3500	0.4878	0.4707	0.3599	0.5726
Observations	7044	7044	7044	7044	7044

Source: Authors' calculations.

Note: OLS model weighted by CZ population in 2008. Sample: commuting zones with at least 25,000 residents in 2008. Year fixed effects and state times period fixed effects included. Additional control variables (interacted with D2005–2007, D2009–2012, and D2013–2016 dummies) same as table 3. Standard errors clustered by state in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5/12/22 12:11 PM

of low-skilled fossil fuel workers are often a decent match for jobs in the green economy. That these manual labor jobs are permanent offers hope that the larger green stimulus discussed on both sides of the Atlantic will provide new employment opportunities as part of a green energy transition. But because job training requirements are higher for green energy jobs than for comparable positions, any green spending plan intending to create green jobs should include funds for job training to ensure a smooth transition into green jobs for displaced workers in fossil fuel and energy intensive sectors.

A second key finding is that, whether due to administrative delays, skill gaps, or the time needed to build new green infrastructure, ARRA's green investments created jobs more slowly than other ARRA investments. Administrative delays such as buy American guidelines, determining prevailing wages to comply with the Davis-Bacon Act, and complying with local regulations may have slowed the initial impact of spending. For instance, less than one-half of DOE funds allocated had been spent by 2011 (Carley, Nicholson-Crotty, and Fisher 2015; Carley 2016). However, administrative delays seem unlikely to explain the permanence of green jobs created by ARRA. Another potential explanation is that federal investments attracted additional private investments in green sectors (Mundaca and Richter 2015) and generally crowded-in state spending (Leduc and Wilson 2017). Many ARRA programs required matching funds from the private sector, and this was particularly true of DOE projects (Council of Economic Advisers 2010). Transforming to a greener economy was expected to support long-term economic growth (Aldy 2013).[34] Our results on the mediating effect of green skills suggest another possibility, as the importance of skills suggests matching funding to skills may affect the time profile of the green ARRA effect.

Such delays are a feature of other public investments that have an important infrastructure component (Ramey 2021). Garin (2019) finds that the job creation effect of construction infrastructure peaked in 2013, but then started declining until going to zero in 2015. Importantly, while our year-by-year results show that most job creation began in 2012, the effect looks permanent for the case of green ARRA spending, especially so in communities with the appropriate set of skills. Given the permanent nature of

34. For example, the DOE's smart grid program invested $4.5 billion in new smart grid technology, which was matched by $6 billion in private sector funds (Council of Economic Advisers 2010). It is reasonable to expect such new infrastructure investment to provide lasting benefits for green employment.

15985-01a-Popp-4thPgs.indd 38 5/12/22 12:11 PM

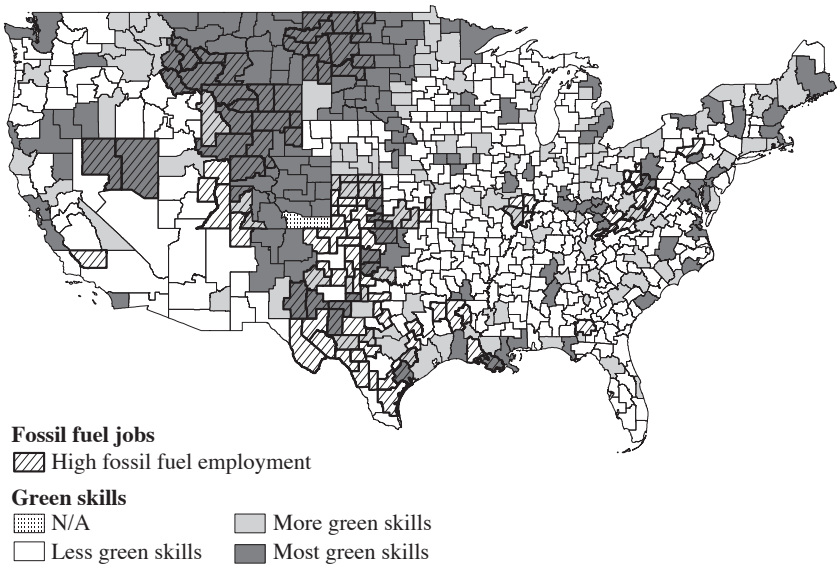
many of the jobs created, green investments can help meet long-run policy goals, such as rebuilding after the pandemic and, more importantly, mitigate the impact of the green energy transition for fossil fuel workers and communities.

Recent proposals for green investments in the United States provide examples of how green fiscal stimuli can serve both roles mentioned above. The Biden administration's $2 trillion American Jobs Plan (AJP) unveiled in spring 2021 includes about $550 billion in green investments (White House 2021a). The bipartisan compromise reached on July 28, 2021, reduces the total investment to just $550 billion but maintains over $200 billion for green investments (White House 2021b). The AJP includes a long-term focus, claiming "this is the moment to reimagine and rebuild a new economy" and promising "to meet the great challenges of our time," including the climate crisis (White House 2021a). To achieve this goal, the AJP includes projects to help workers whose jobs may be at risk as the world transitions away from fossil fuels. Green investments are more appropriate as part of such long-term planning addressing climate change than as short-term economic stimulus. Table A3 of online appendix A1 highlights major areas of green spending in each plan. Examples of green investments include plugging orphan oil and gas wells and cleaning abandoned mines to create jobs for displaced energy workers, developing charging infrastructure for electric vehicles, and improving water infrastructure, such as by replacing lead pipes. While such jobs offer opportunities for displaced energy workers, they also illustrate the challenges of place-based policies to aid distressed energy communities, as jobs such as plugging wells or cleaning abandoned mines are unlikely to be permanent jobs.

Our third key finding is that workers must have the skills needed in green jobs for green fiscal stimuli to be successful. Thus, green spending policies may not work equally well in distressed communities lacking the appropriate green skills. Recall as well that the availability of green skills increases the short-run effectiveness of green spending more than the long-run effectiveness. In the short run, this means that a good match between required skills and local skill endowments is essential for successful fiscal stimuli. In the long-run, distressed energy communities may need to combine green spending with appropriate retraining policies.

Moreover, green investments may create spatial inequities that affect political acceptability and the potential for different regions to benefit from the transition to a greener economy. We draw attention to two dimensions of such inequalities here. Both relate to the geographic distribution of green skills.

Figure 5. Geographic Distribution of Green Skills



Source: Authors' calculations.

Note: Figure shows commuting zones in the top quartile (dark), second quartile (light), and bottom 50th percentile of green skills, using our measure of share of employment in jobs in the top quartile of green skill requirement based on 2019 data. Communities in the top 20 percent of the share of fossil fuel employment are outlined and highlighted with dark stripes. Census data on green skills are missing for two commuting zones in Colorado.

First, it is obvious that communities with a higher share of fossil fuel jobs will experience large negative shocks due to the reduction in demand for fossil fuels. In the United States, many communities dependent on coal have already experienced economic decline, as both lower natural gas prices and the expansion of wind energy reduced demand for coal (Fell and Kaffine 2018; Weber 2020). More stringent emission reduction goals will eventually bring similar declines to communities where oil and gas drilling is prominent. Our results suggest that not all communities with a high share of fossil fuel jobs possess the right engineering and technical skills to attract green activities. Figure 5 illustrates the overlap between the presence of fossil fuel jobs (dark stripes indicating commuting zones in the top 20 percent of the share of fossil fuel employment) and green skill intensity in 2019 (shaded). There is a large heterogeneity in the level of green competences in fossil fuel intensive communities. Areas in the West and Midwest appear well prepared for the low-carbon transition.

Many communities in both Wyoming and North Dakota have high levels of green skills. Although beyond the scope of this paper, that may in part be due to the abundance of wind energy resources in these regions. While there is larger variation in green skills endowments in the fossil fuel intensive regions in the South, these regions mostly specialize in oil and gas, which will still play a role in the transition to a greener energy economy. In contrast, the Appalachian region is facing dramatic decreases in demand for coal and also has several commuting zones with low levels of green skills.

Second, communities with a higher share of green skills are also wealthier, as shown in table B13 of the online appendix. Thus, using large green stimuli as part of a green energy transition has the potential to exacerbate regional inequities. While the AJP attempts to address regional inequities by focusing infrastructure investments such as water infrastructure on disadvantaged communities, communities with the appropriate level of green competences will attract complementary private-sector investments in green enabling sectors, such as producers of wind turbines or electric vehicles, that are generally high-tech and concentrated in wealthier regions (Bontadini and Vona 2020). This may conjure a trade-off between choosing to specialize in the production of green technologies and using green spending to create new opportunities for distressed communities, especially in regions such as Appalachia.

Previous literature on place-based policies shows that investments in vocational and on-the-job training can be particularly effective in distressed regions (Bartik 2020), reinforcing our claim that well-targeted job training investments should be a key part of green fiscal plans to come. Worker-level evidence on the effectiveness of training investments is more mixed, but the recent paper by Hyman (2018) provides somewhat encouraging results on the effect of Trade Adjustment Assistance on earnings and reemployment of displaced workers. However, the lessons of previous spatially localized shocks—that is, the China shock and the decline of coal communities (Autor, Dorn, and Hanson 2021)—indicate that the labor market adjustment can be extremely long and painful for affected workers. It is thus unlikely that in the context of localized shocks, such as those to coal or shale gas communities, training investments alone can be a panacea. Future research should consider the potential of broader policies not just to help those directly affected by energy sector job losses but also to tackle the spillover effects to the rest of the local economy when these jobs move out.

VII. Conclusion

We perform a comprehensive evaluation of the economic effect of green stimulus using the historical experience of the American Recovery and Reinvestment Act, which represents the largest push to the green economy to date. Our results inform both current policy debates and address longer-term concerns about job losses in the transition to a green economy. Currently, Green New Deal programs are seen by some policy advocates as a win-win solution both to relaunch sluggish economic growth in developed countries and to tackle climate change. While the size of the green stimulus of 2009 was small compared to what is proposed as part of a post-COVID-19 recovery, our research highlights interesting features of a green stimulus that can offer guidance to the design of future green stimulus programs, as well as to the broader potential of green investments to ease the employment shocks from a green energy transition.

First, our results suggest green ARRA works more slowly than other stimulus investments. The effect of green ARRA on total employment is ambiguous; while the gross effect is twenty-five jobs per year created per $1 million of green ARRA, the presence of pre-trends makes it difficult to discern whether these are new jobs or simply a return to pre–Great Recession levels of employment. However, what is clear is that both the changes in total employment and those in specific sectors persist over time. The persistency of the job creation effect is clearly a positive aspect of the green fiscal stimulus and a notable difference from previous studies of other ARRA investments, which generally find short-term effects. In contrast, we do not find evidence of short-run employment gains for green ARRA. Thus, green stimulus investments appear more effective for reshaping an economy than for restarting an economy. While our focus is on the potential employment benefits from green investments, future research should also consider the potential environmental benefits of green stimulus, as the long-run impacts on employment suggest that green investments lead to durable changes in the green economy. However, since these investments do not come with regulatory requirements to reduce emissions, do these long-run changes lead to an improved environment?

Second, the impact of the green stimulus becomes much clearer when we explore several dimensions of heterogeneity. Green ARRA creates more jobs in commuting zones with larger initial shares of occupations that use intensively green skills. In particular, $1 million of green ARRA spending creates approximately 1.5 as many jobs in areas in the top quartile of the green skills distribution than in the median commuting zone. Our

descriptive evidence suggests that the skill gap between green energy and fossil fuel jobs is modest, but green jobs require significantly more on-the-job training than fossil fuel jobs. Moreover, the geographical distribution of green skills may complicate the reallocation process required by the energy transition. On the one hand, the geographic concentration of these skills among fossil fuel dependent communities varies. On the other hand, communities with a higher share of green skills are also wealthier, so that green investments potentially enhance opportunities in communities already in position to support a green economy. Additional investments in vocational and on-the-job training could improve the effectiveness of green stimuli in regions without the required green skills. Evaluation of such training programs is left for future work.

Third, a green stimulus has potential to reshape an economy and thus may have important distributional effects. Green ARRA especially increases the demand for manual laborers. Beyond the direct impacts of a green stimulus, these results also have broader implications for whether governments can help ease labor market transitions in response to environmental policy using place-based policies. Recent studies suggest that environmental regulation may reduce jobs in specific sectors, particularly for manual (Marin and Vona 2019) and unskilled labor (Yip 2018). In contrast, subsidies to green infrastructure can benefit unskilled workers and thus may enhance the political support for other climate policies. However, wage gains did not follow the increase in the demand for manual tasks in areas receiving higher green subsidies. Exploring whether this is due to the fact that green jobs in construction are of low quality compared to similar jobs or to the widespread deterioration of the bargaining power of unskilled workers requires the use of longitudinal worker-level data and is left for future research.

ACKNOWLEDGMENTS This project has been supported in part through the Smart Prosperity Institute Research Network and its Greening Growth Partnership, which is supported by a Social Sciences and Humanities Research Council of Canada Partnership Grant (no. 895-2017-1018), as well as by Environment and Climate Change Canada's Economics and Environmental Policy Research Network (EEPRN). This work was also supported by Horizon 2020 Framework Programme, project INNOPATHS (grant no. 730403). We thank Joe Aldy, James Stock, Gabriel Chodorow-Reich, Valerie Ramey, Carolyn Fischer, Michelle Li, and Michael Chen for useful suggestions and discussions, especially on green ARRA data. For helpful comments, we thank seminar

participants at Harvard University, Duke Kunshan University, the London School of Economics, University of Bremen, SPRU at University of Sussex, University of Newcastle, APPAM, CMCC Ca' Foscari, CESifo, the ETH workshop Energy, Innovation and Growth: Theory and Empirics, the Coalition of Finance Ministers for Climate Action workshop on The Political Economy of Carbon Pricing Post Covid-19, the workshop Where and How to Invest the EU Recovery Funds? at Universidad Pontificia Comillas of Madrid, the Resources for the Future workshop Evaluating Climate-Oriented Economic Recovery Programs, and the Greening Growth Partnership and Economics and Environmental Policy Research Network Annual Symposium.

15985-01a-Popp-4thPgs.indd 44 5/12/22 12:11 PM

References

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

Acemoglu, Daron, and Pascual Restrepo. 2020. "Robots and Jobs: Evidence from US Labor Markets." *Journal of Political Economy* 128, no. 6: 2188–244.

Agrawala, Shardul, Damien Dussaux, and Norbert Monti. 2020. "What Policies for Greening the Crisis Response and Economic Recovery? Lessons Learned from Past Green Stimulus Measures and Implications for the COVID-19 Crisis." Working Paper 164. Paris: Organization for Economic Cooperation and Development. https://www.oecd-ilibrary.org/environment/what-policies-for-greening-the-crisis-response-and-economic-recovery_c50f186f-en.

Aldy, Joseph E. 2013. "Policy Monitor—A Preliminary Assessment of the American Recovery and Reinvestment Act's Clean Energy Package." *Review of Environmental Economics and Policy* 7, no. 1: 136–55.

Aldy, Joseph E., and William A. Pizer. 2015. "The Competitiveness Impacts of Climate Change Mitigation Policies." *Journal of the Association of Environmental and Resource Economists* 2, no. 4: 565–95.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113, no. 1: 151–84.

American Recovery and Reinvestment Act (ARRA) of 2009, Pub. L. 111-5, 123 Stat. 115, section 3.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, N.J.: Princeton University Press.

Auerbach, Alan J., Yuriy Gorodnichenko, and Daniel Murphy. 2020. "Local Fiscal Multipliers and Fiscal Spillovers in the USA." *IMF Economic Review* 68, no. 1: 195–229.

Autor, David, David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103, no. 6: 2121–68.

Autor, David, David Dorn, and Gordon H. Hanson. 2021. "On the Persistence of the China Shock." In the present volume of *Brookings Papers on Economic Activity*.

Autor, David, David Dorn, Gordon Hanson, and Kaveh Majlesi. 2020. "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure." *American Economic Review* 110, no. 10: 3139–83.

Autor, David, Frank Levy, and Richard Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118, no. 4: 1279–333.

Bartik, Timothy J. 2020. "Using Place-Based Jobs Policies to Help Distressed Communities." *Journal of Economic Perspectives* 34, no. 3: 99–127.

Bontadini, Filippo, and Francesco Vona. 2020. "Anatomy of Green Specialization: Evidence from EU Production Data, 1995–2015." Working paper 21/2020. Paris: Science Po OFCE. https://www.ofce.sciences-po.fr/pdf/dtravail/OFCEWP2020-21.pdf.

Carley, Sanya. 2016. "Energy Programs of the American Recovery and Reinvestment Act of 2009." *Review of Policy Research* 33, no. 2: 201–23.

Carley, Sanya, Sean Nicholson-Crotty, and Eric J. Fisher. 2015. "Capacity, Guidance and the Implementation of the American Recovery and Reinvestment Act." *Public Administration Review* 75, no. 1: 113–25.

Castellanos, Kenneth A., and Gareth Heutel. 2019. "Unemployment, Labor Mobility, and Climate Policy." Working paper 25797. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25797.

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. 2020. "Regional Data in Macroeconomics: Some Advice for Practitioners." *Journal of Economic Dynamics and Control* 115:103875.

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.

Conley, Timothy G., and Bill Dupor. 2013. "The American Recovery and Reinvestment Act: Solely a Government Jobs Program?" *Journal of Monetary Economics* 60, no. 5: 535–49.

Council of Economic Advisers. 2010. *The Economic Impact of the American Recovery and Reinvestment Act of 2009, Fourth Quarterly Report*. Washington: Council of Economic Advisers. https://obamawhitehouse.archives.gov/administration/eop/cea/factsheets-reports/economic-impact-arra-4th-quarterly-report/summary.

Council of Economic Advisers. 2013. *The Economic Impact of the American Recovery and Reinvestment Act of 2009, Ninth Quarterly Report*. Washington: Council of Economic Advisers. https://obamawhitehouse.archives.gov/sites/default/files/docs/cea_9th_arra_report_final_pdf.pdf.

Council of Economic Advisers. 2014. "The Economic Impact of the American Recovery and Reinvestment Act Five Years Later." In *Economic Report of the President*. Washington: United States Government Printing Office.

Cullen, Joseph A., and Erin T. Mansur. 2017. "Inferring Carbon Abatement Costs in Electricity Markets: A Revealed Preference Approach Using the Shale Revolution." *American Economic Journal: Economic Policy* 9, no. 3: 106–33.

Dupor, Bill, and Peter B. McCrory. 2018. "A Cup Runneth Over: Fiscal Policy Spillovers from the 2009 Recovery Act." *Economic Journal* 128, no. 611: 1476–508.

Dupor, Bill, and M. Salif Mehkari. 2016. "The 2009 Recovery Act: Stimulus at the Extensive and Intensive Labor Margins." *European Economic Review* 85: 208–28.

Environmental Protection Agency. 2009. American Recovery and Reinvestment Act Funding for the National Clean Diesel Funding Assistance Program: Request for Applications. Washington: Environmental Protection Agency. https://www.epa.gov/sites/default/files/2015-10/documents/recovery-act-national-clean-diesel-rfa.pdf.

Fell, Harrison, and Daniel T. Kaffine. 2018. "The Fall of Coal: Joint Impacts of Fuel Prices and Renewables on Generation and Emissions." *American Economic Journal: Economic Policy* 10, no. 1: 90–116.

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram. 2018. "Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program." *Quarterly Journal of Economics* 133, no. 3: 1597–644.

Gabe, Todd M., and Jaison R. Abel. 2012. "Specialized Knowledge and the Geographic Concentration of Occupations." *Journal of Economic Geography* 12, no. 2: 435–53.

Garin, Andrew. 2019. "Putting America to Work, Where? Evidence on the Effectiveness of Infrastructure Construction as a Locally Targeted Employment Policy." *Journal of Urban Economics* 111:108–31.

Gathmann, Christina, and Uta Schönberg. 2010. "How General Is Human Capital? A Task-Based Approach." *Journal of Labor Economics* 28, no. 1: 1–50.

Greenstone, Michael. 2002. "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures." *Journal of Political Economy* 110, no. 6: 1175–219.

Hafstead, Marc A., and Roberton C. Williams III. 2018. "Unemployment and Environmental Regulation in General Equilibrium." *Journal of Public Economics* 160:50–65.

Helm, Dieter. 2020. "The Environmental Impacts of the Coronavirus." *Environmental and Resource Economics* 76, no. 1: 21–38.

Hyman, Benjamin G. 2018. "Can Displaced Labor Be Retrained? Evidence from Quasi-Random Assignment to Trade Adjustment Assistance." Working Paper. Philadelphia: Mack Institute for Innovation Managment. https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2018/11/FP0385_WP_2018Nov.pdf.

Imbens, Guido, and Joshua Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62, no. 2: 467–75.

Jaeger, David A., Joakim Ruist, and Jan Stuhler. 2018. "Shift-Share Instruments and the Impact of Immigration." Working Paper 24285. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24285.

Kahn, Matthew, and Erin Mansur. 2013. "Do Local Energy Prices and Regulation Affect the Geographic Concentration of Employment?" *Journal of Public Economics* 101:105–14.

Kambourov, Gueorgui, and Iourii Manovskii. 2009. "Occupational Specificity of Human Capital." *International Economic Review* 50, no. 1: 63–115.

Konisky, David M., and Sanya Carley. 2021. "What We Can Learn from the Green New Deal about the Importance of Equity in National Climate Policy." *Journal of Policy Analysis and Management* 40, no. 3: 996–1002.

Leduc, Sylvain, and Daniel J. Wilson. 2017. "Are State Governments Roadblocks to Federal Stimulus? Evidence on the Flypaper Effect of Highway Grants in the 2009 Recovery Act." *American Economic Journal: Economic Policy* 9, no. 2: 253–92.

Marin, Giovanni, and Francesco Vona. 2019. "Climate Policies and Skill-Biased Employment Dynamics: Evidence from EU Countries." *Journal of Environmental Economics and Management* 98:102253.

Marin, Giovanni, and Francesco Vona. 2021. "The Impact of Energy Prices on Socioeconomic and Environmental Performance: Evidence from French Manufacturing Establishments, 1997–2015." *European Economic Review* 135:103739.

Metcalf, Gilbert E., and James H. Stock. 2020. "Measuring the Macroeconomic Impact of Carbon Taxes." In *American Economic Association Papers and Proceedings*, vol. 110. Pittsburgh, Pa.: AEA Publications.

Moretti, Enrico. 2010. "Local Multipliers." *American Economic Review* 100, no. 2: 373–77.

Morgenstern, Richard, William Pizer, and Jhih-Shyang Shih. 2002. "Jobs versus the Environment: An Industry-Level Perspective." *Journal of Environmental Economics and Management* 43, no. 3: 412–36.

Mundaca, Luis, and Jessika Luth Richter. 2015. "Assessing 'Green Energy Economy' Stimulus Packages: Evidence from the U.S. Programs Targeting Renewable Energy." *Renewable and Sustainable Energy Reviews* 42:1174–86.

Muro, Mark, Adie Tomer, Ranjitha Shivaram, and Joseph Kane. 2019. "Advancing Inclusion through Clean Energy Jobs." Washington: Brookings Institution, Metropolitan Policy Program. https://www.brookings.edu/research/advancing-inclusion-through-clean-energy-jobs/.

Nakamura, Emi, and Jón Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review* 104, no. 3: 753–92.

Ramey, Valerie A. 2021. "The Macroeconomic Consequences of Infrastructure Investment." In *Economic Analysis and Infrastructure Investment*, edited by Edward L. Glaeser and James M. Poterba. Chicago: University of Chicago Press.

Tomer, Adie, Joseph W. Kane, and Caroline George. 2021. "How Renewable Energy Jobs Can Uplift Fossil Fuel Communities and Remake Climate Politics." Washington: Brookings Institution, Metropolitan Policy Program. https://www.brookings.edu/research/how-renewable-energy-jobs-can-uplift-fossil-fuel-communities-and-remake-climate-politics/.

US Department of Energy. 2010. "Recovery Act State Memos: Massachusetts." Washington: Department of Energy. https://www.energy.gov/downloads/massachusetts-recovery-act-state-memo.

US Office of the Vice President. 2010. "100 Recovery Act Projects That Are Changing America." Washington: Environmental Protection Agency. https://archive.epa.gov/recovery/web/pdf/100-recovery-act-projects-changing-america-report.pdf.

Vona, Francesco. 2019. "Job Losses and the Political Acceptability of Climate Policies: Why the 'Job-Killing' Argument Is So Persistent and How to Overturn It." *Climate Policy* 19, no. 4: 524–32.

Vona, Francesco, Giovanni Marin, and Davide Consoli. 2019. "Measures, Drivers and Effects of Green Employment: Evidence from US Metropolitan and Non-Metropolitan Areas, 2006–2014." *Journal of Economic Geography* 19, no. 5: 1021–48.

Vona, Francesco, Giovanni Marin, Davide Consoli, and David Popp. 2018. "Environmental Regulation and Green Skills: An Empirical Exploration." *Journal of the Association of Environmental and Resource Economists* 5, no. 4: 713–53.

Walker, W. Reed. 2013. "The Transitional Costs of Sectoral Reallocation: Evidence from the Clean Air Act and the Workforce." *Quarterly Journal of Economics* 128:1787–835.

Weber, Jeremy G. 2020. "How Should We Think about Environmental Policy and Jobs? An Analogy with Trade Policy and an Illustration from US Coal Mining." *Review of Environmental Economics and Policy* 14, no. 1: 44–66.

White House. 2021a. "Fact Sheet: The American Jobs Plan." March 31. https://www.whitehouse.gov/briefing-room/statements-releases/2021/03/31/fact-sheet-the-american-jobs-plan/.

White House. 2021b. "Fact Sheet: Historic Bipartisan Infrastructure Deal." July 28. https://www.whitehouse.gov/briefing-room/statements-releases/2021/07/28/fact-sheet-historic-bipartisan-infrastructure-deal/.

Wilson, Daniel J. 2012. "Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act." *American Economic Journal: Economic Policy* 4, no. 3: 251–82.

Yip, Chi Man. 2018. "On the Labor Market Consequences of Environmental Taxes." *Journal of Environmental Economics and Management* 89:136–52.

Yip, Chi Man. 2020. "The Ins and Outs of Employment: Labor Market Adjustments to Carbon Taxes." Working Paper. https://arts.ucalgary.ca/sites/default/files/teams/27/chi-man-yip-jmp-ins-and-out-employment-labor-market-adjustments-carbon-taxescompressed.pdf.

Comments and Discussion

COMMENT BY

GABRIEL CHODOROW-REICH Popp, Vona, Marin, and Chen have written an ambitious paper on an important topic. Advocates of a Green New Deal sometimes claim a substantial macroeconomic dividend from the investments required to transition the US economy away from fossil fuels. The authors provide an empirical assessment of this possibility, taking account of both the workers likely to be displaced in such a transition and the evidence from the clean energy investments contained in the 2009 American Recovery and Reinvestment Act (ARRA). They arrive at a more nuanced conclusion, yet retain optimism for the potential for clean energy investments to absorb displaced workers.

I will divide my comment into three parts. The first concerns the relationship among temporary stimulus, permanent job creation, and permanent structural transformation and what one can learn from the ARRA for the paper's ultimate question. The second articulates my view of the difficult econometric and measurement challenge the authors face in studying the green aspects of the ARRA. The third broadens to consider lessons from other structural transformations.

TEMPORARY STIMULUS VERSUS PERMANENT STRUCTURAL TRANSFORMATION

My first point concerns the relationship among temporary stimulus, permanent job creation, and permanent structural transformation. There are two main reasons why policymakers might consider the green transformation as part of macroeconomic policy. The first is an opportunity. Governments can pull forward green investments to periods when resources are otherwise idle, namely, a recession. This is properly termed fiscal stimulus. According to standard economic theory, by raising the aggregate demand for labor, temporary fiscal stimulus can increase total employment temporarily while resources remain idle. Targeted investments can lead to

15985-01b-Popp_Comm&GD-4thPgs.indd 50 5/12/22 12:10 PM

Table 1. DOE and EPA Programs

Agency	*Program*	*Appropriation ($ billions)*	*Outlays as of 2013:Q4 ($ billions)*
DOE	Energy efficiency and renewable energy	16.80	15.50
EPA	State and tribal assistance grants	6.34	6.21
DOE	Defense environmental cleanup	5.13	5.11
DOE	Electricity delivery and energy reliability	4.50	3.97
DOE	Fossil Energy Research and Development	3.40	1.22
DOE	Western Area Power Administration Borrowing Authority	3.25	0.22
DOE	Bonneville Power Administration fund	3.25	1.36
DOE	Title 17 Incentives for Innovative Technologies loan guarantee	2.50	1.33
DOE	Science	1.60	1.59
EPA	Hazardous Substances Superfund	0.60	0.62
DOE	Non-defense environmental cleanup	0.48	0.44
DOE	Energy Transformation Acceleration Fund	0.40	0.35
DOE	Uranium Enrichment Decontamination and Decommissioning Fund	0.39	0.39
EPA	Leaking Underground Storage Tank Trust Fund	0.20	0.19
EPA	Environmental Program and Management	0.06	0.08
DOE	Advanced Technology Vehicles Manufacturing Loan Program	0.01	0.01
DOE	Construction, rehabilitation, operation, and maintenance, Western Area Power Administration	0.01	0.01

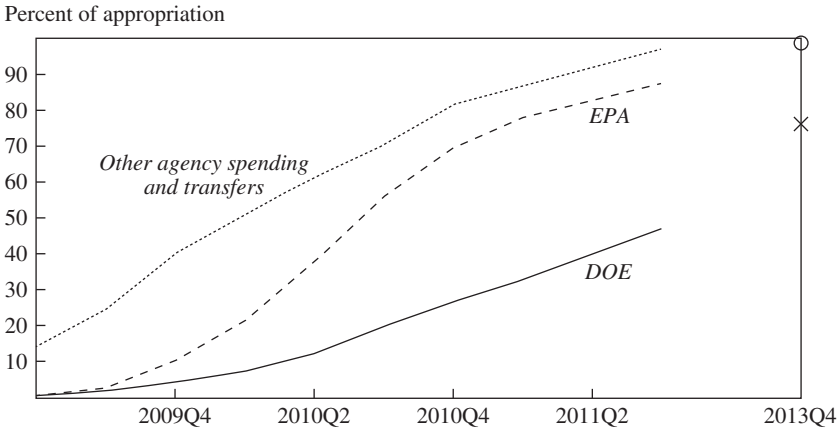
Sources: Agency FAR reports and author's calculations.

permanent reallocation of employment toward green industries or to certain geographic areas but do not permanently increase aggregate labor demand.

The second reason policymakers might think of a green transformation in macroeconomic terms stems from a problem. Many workers currently employed in industries that emit greenhouse gas (GHG) eventually will have to find some other way to make a living. Because of frictions involved in finding new employment or gaining new skills, policy may want to provide support for the losers in the industrial reshuffling. Temporary stimulus cannot satisfy this objective, and measures to address it should not be conflated with temporary stimulus. Instead, such measures involve sustained demand-side support such as investment incentives for new businesses to locate in heavily impacted areas, supply-side interventions such as worker retraining, or simply financial support for displaced workers such as wage insurance or pension payments.

What type of spending did the ARRA contain? Table 1 lists the Department of Energy (DOE) and Environmental Protection Agency (EPA)

Figure 1. Time Path of ARRA Spending by Agency



Sources: Agency FAR reports and author's calculations.
Note: Spending in 2011:Q4–2013:Q3 omitted for lack of data.

spending programs along with their appropriation and total outlays through 2013:Q4. Most of these programs fall under the umbrella of temporary stimulus, such as financing energy efficiency retrofits or investments in the smart grid. Almost all of the spending had concluded by the end of 2013.[1]

Figure 1 shows the time path of the spend-out of various components of the ARRA, expressed as a share of their appropriation.[2] The dotted line shows the path of all non-energy outlays. Most non-energy outlays had concluded by the end of 2010. The dashed line shows EPA spending, which accounts for $7.2 billion of the combined (EPA and DOE) $49 billion appropriation. EPA spending ramps up more slowly than total ARRA but at a pace similar to most of the non-transfer components, and about three-quarters had been outlaid by the end of 2010. In contrast, the Department of Energy had outlaid only about one-quarter of its total appropriation by the end of 2010. The X shows the total share in 2013:Q4. DOE outlays

1. Much of the difference between outlays as of 2013:Q4 and the initial appropriation involves programs that never spent out their full appropriation. Such programs include a combined $6.5 billion in borrowing authority for the Bonneville Power Administration and Western Area Power Administration, of which only $1.58 billion had been used as of 2013:Q4, and the Title 17 innovative technologies loan guarantee program, which had a $2.5 billion appropriation but recorded charges of only $1.3 billion as of 2013:Q4.

2. The figure omits spending in 2011:Q4–2013:Q3 as I do not have data for those quarters and the original repositories of the data are no longer available.

15985-01b-Popp_Comm&GD-4thPgs.indd 52 5/12/22 12:10 PM

were larger in 2011–2012 than they were in 2009–2010. Nonetheless, most spending had been completed by the end of 2013.

This discussion helps to frame the paper's headline finding of delayed but possibly permanent employment effects from the clean energy provisions of the ARRA. The slow spend-out of DOE programs explains why they had little impact in 2009–2010. Even so, the economy remained demand-constrained in 2011 and 2012, as evidenced, for example, by the Federal Reserve's target federal funds rate remaining close to zero until the end of 2015. Thus, it seems plausible that clean energy investments in that period absorbed slack resources and resulted in higher total employment. It would be more surprising for such spending to continue to produce meaningfully higher economy-wide employment in 2019, well after all of the spending took place and when the economy was at or near full employment. For this reason, I am inclined to treat the paper's wide range of possible estimates of long-run employment effects as providing little reason to move my prior that such effects would be small. Of course, it could be the case that certain geographic areas continued to benefit because of earlier investments in business capital, which highlights the distinction between local and national multipliers.[3]

ECONOMETRIC AND MEASUREMENT CHALLENGE My second point concerns the econometric challenge of estimating the employment effects of the ARRA's clean energy provisions. As discussed carefully in the paper, ARRA spending on green programs was likely targeted toward areas with existing green infrastructure or potential. For example, research and development grants went to places with scientific laboratories already working on energy matters and wind turbines were built in places with high wind potential.

The nonrandom allocation of funds raises the possibility of omitted variable bias. The generic concern is that areas with existing green infrastructure share other characteristics that made them subject to correlated shocks or have similar loadings on aggregate shocks, such as less exposure to the 2008 recession. For example, maybe these places also had a less dramatic house price cycle. This is a failure of the law of large numbers (Borusyak, Hull, and Jaravel 2022)—with only one ARRA to study, it is

3. In theory, an increase in investment in an area could permanently pull in workers from other areas, making the local employment effects larger than the national effects (Chodorow-Reich 2019). Even so, the geographic concentration of fossil fuel–related jobs suggests that the local multiplier concept analyzed in the paper may be especially informative, since the objective is in part to reduce the negative effect on highly impacted local communities.

possible that other forces happened to have an impact in these particular areas at the same time. Indeed, the presence of pre-trends indicates *something* about these areas that made them correlated even before they ever received ARRA funds.

The specific concern is that the economy was becoming greener even absent the ARRA, resulting in faster growth in areas with existing green infrastructure or potential. Pre-trends are an imperfect diagnostic for this concern, since secular greening may have been more important post-2008 than before. Importantly, this issue is quite different from the typical omitted variable problem with the ARRA or other fiscal stimulus, which is that funds were targeted to harder hit areas, biasing down the estimated jobs effect. Here the direction is the opposite—funds may have gone to areas already greening, and the additional green jobs were due to the secular transition rather than the ARRA per se.

An example can illustrate. There were large ARRA grants and tax credits to support wind and solar development and installation. These provisions obviously were most likely to benefit areas with lots of wind and sun. In many cases, the physical environment provides excellent variation for identifying causal effects—think, for example, of papers using rainfall shocks as an excluded instrument for income or physical land constraints as an excluded instrument for housing supply. But in the context of the ARRA, a problem arises if wind and solar were becoming more competitive anyway due to the rapidly declining cost of imported solar panels and wind turbines, because the declining cost would generate wind and solar investment in places with high wind and sun even absent the ARRA. This problem explains why the authors choose to control for wind and sun potential rather than exploiting it as a physical determinant of green ARRA allocation. Put this way, however, it is immediately clear that the same concern applies to any shift-share design.[4]

Taking stock, one can understand why the econometric specification contains many controls, yet this makes it difficult to assess what determines the remaining variation in green ARRA allocation. The paper emphasizes the absence of pre-trends in the results for green and occupational employment in arguing for a cleaner interpretation of these results, but ultimately

4. The issue is a bit more subtle in the interaction specification but does not disappear. In particular, if green suitability is measured with error, then the interaction coefficient may also pick up the tendency for funds to go to areas expected to expand based on unobservables to the econometrician.

15985-01b-Popp_Comm&GD-4thPgs.indd 54 5/12/22 12:10 PM

Table 2. Measuring Green Jobs

	%
Population-weighted median green share (GS)	3.05
Bootstrap-weighted median SD	0.16
Weighted median coefficient of variation	5.24

Sources: 2013 American Community Survey and author's calculations.

macroeconomic policy must care about total employment, not sectoral employment. And the stubborn endurance of pre-trends in the results for total employment suggests there still may be common shocks or trends affecting these areas.

A final challenge to estimating the impact on green jobs arises even with a perfect research design. The paper defines greenness by occupation rather than industry. Measuring green jobs therefore requires survey data with occupation codes, which the paper takes from the American Community Survey (ACS). But green jobs are a small share of the total economy, about 3 percent of total hours worked. Dividing the ACS into commuting zones and then restricting to the 3 percent of hours worked in green occupations results in pretty small sample sizes of green workers. Table 2 reports the variation in green share due only to sampling variability. To compute this table, I drew fifty bootstrap samples of individuals by public use microdata area (PUMA, i.e., a stratified bootstrap sample following the ACS sample design) in the 2013 ACS, computed the green share following the paper's occupational categorization, and aggregated the bootstrap samples up to the commuting zone level. The weighted median coefficient of variation across commuting zones is 5.24 percent. This implies a standard deviation of roughly $100 \times \sqrt{2 \times 0.0524^2} = 7$ percentage points in the growth rate of green jobs in the typical commuting zone due only to sampling variability. Of course, the standard errors already reflect this sampling variability; my point is simply to highlight the difficulty in this setting of generating precise multiplier estimates for green jobs.

LESSONS FROM OTHER STRUCTURAL TRANSFORMATIONS My final point is in regard to the broader aspects of the clean energy transformation. The first part of the paper provides a very nice analysis of the skill gap between GHG-emitting industries and green industries, motivated by lessons from earlier episodes of structural transformation. This important analysis should somewhat reassure policymakers, though I will note that the general equilibrium reallocation of workers need not necessarily involve the same people who leave GHG-emitting industries to then find work in green industries.

What we know about the secular decline of major industrial sectors offers little overall encouragement. Displaced workers often do not quickly transition to new jobs and instead suffer harmful long-term consequences, as illustrated for example by Autor, Dorn, and Hanson (2021). Job retraining may be part of the solution (Hyman 2018) but is particularly difficult for older, less educated workers. Other proposals such as wage insurance seem worth considering but go considerably beyond the scope of my remarks here. Instead, I will reemphasize my first point about the distinction between temporary stimulus and long-term solutions. Some of the proposed investments discussed in the paper, such as plugging orphan oil wells or cleaning abandoned mines, might qualify as temporary stimulus if the US economy were to enter a demand-driven recession, and surely these projects are worthwhile for their environmental benefits, but these sorts of onetime investments or cleanup projects cannot provide permanent alternative work for workers displaced by the clean energy transformation.

On the other hand, a strong macroeconomy is conducive to sectoral transformation. Johannes Wieland and I refer to this as macroeconomic policy "greasing the wheels" of reallocation by accommodating higher inflation (Chodorow-Reich and Wieland 2020). Guerrieri and others (2021) explicitly develop this point. With downward wage rigidity binding in a recession, wages cannot fall in shrinking industries. Without this price signal for workers to reallocate, they instead remain looking for work in declining industries, slowing down the transformation and exacerbating unemployment. This gives a rationale for wanting a hot economy to facilitate a transition and even to tolerate possibly above-trend inflation.

CONCLUSION This is a thought-provoking paper on an important and challenging topic. Overall, I find myself somewhat less optimistic than the paper about the labor market challenges of the green transition, although also more optimistic than before I read it.

REFERENCES FOR THE CHODOROW-REICH COMMENT

Autor, David, David Dorn, and Gordon H. Hanson. 2021. "On the Persistence of the China Shock." In the present volume of *Brookings Papers on Economic Activity*.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-Experimental Shift-Share Research Designs." *Review of Economic Studies* 81, no. 1: 181–213.

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, and Johannes Wieland. 2020. "Secular Labor Reallocation and Business Cycles." *Journal of Political Economy* 128, no. 6: 2245–87.

Guerrieri, Veronica, Guido Lorenzoni, Ludwig Straub, and Iván Werning. 2021. "Monetary Policy in Times of Structural Reallocation." In *Economic Policy Symposium Proceedings: Macroeconomic Policy in an Uneven Economy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Hyman, Benjamin G. 2018. "Can Displaced Labor Be Retrained? Evidence from Quasi-Random Assignment to Trade Adjustment Assistance." Working Paper. Philadelphia: Mack Institute for Innovation Managment. https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2018/11/FP0385_WP_2018Nov.pdf.

COMMENT BY

VALERIE A. RAMEY This very interesting paper by David Popp and colleagues offers a comprehensive analysis of the employment effects of the transition from fossil fuels to green energy. A particular focus of the paper is assessing the extent to which green fiscal stimuli can mitigate the negative employment effects on fossil fuel workers. This question is important because government policies designed to convert energy production from fossil fuels to green energy may face political opposition from the potential losers.

The paper consists of four parts. The first part surveys the literature on the effects of environmental policies on employment, with attention to heterogeneity in skills and geography. The second part presents evidence comparing the skill requirements of green jobs with the skill endowment of workers in fossil fuel industries. The depth and breadth of the analysis of the various types of heterogeneity make these two parts a valuable resource for academics and policymakers. The third part analyzes the employment effects of the green spending in the 2009 American Recovery and Reinvestment Act (ARRA) at the commuting zone level. This part of the analysis is an ambitious undertaking since features of the data present multiple challenges to obtaining definitive answers. Finally, the fourth part discusses the policy implications of the findings of the paper.

Most of my comments concern the interpretation of the ARRA estimates. First, I discuss what the authors' local estimates imply about aggregate effects and what macroeconomic theory predicts about the aggregate effects of infrastructure spending. Second, I summarize the authors' green ARRA employment findings and compare them to the estimated effects of the highway spending components of the ARRA. Third, I discuss green incentives that were not included in the authors' analysis, specifically the effects of tax credits for rooftop solar, and how the effects of these incentives may confound the authors' estimates. I also discuss advantages that rooftop solar has over some alternatives. Finally, I return to the authors' motivating question

regarding potential employment losses and consider the broader impacts of the green transition on fossil fuel communities. I illustrate the challenges with the example of the mining towns in Asturias, Spain.

WHAT THEORY PREDICTS ABOUT THE EFFECTS OF INFRASTRUCTURE SPENDING

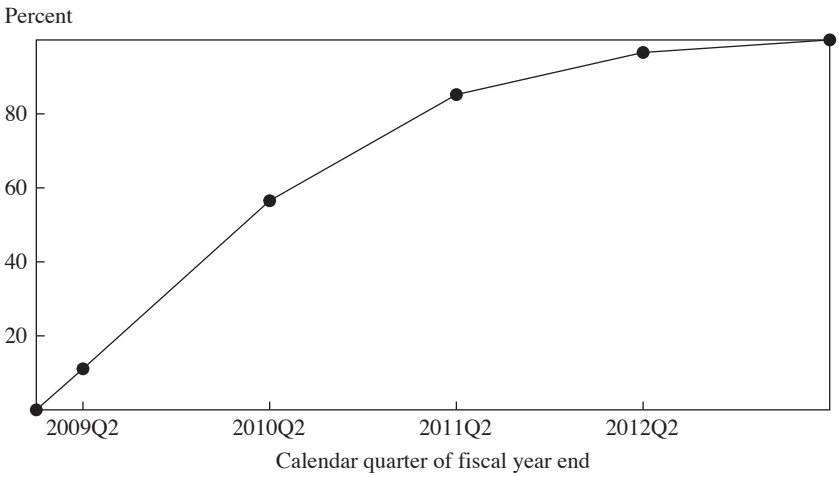
The authors present estimates of the employment effects of the green spending in the ARRA using variations across commuting zones. They consider the effects at a variety of horizons and find small effects in the first few years but then apparently permanent positive effects on job creation. Before using these estimates to guide policy, it is important to put these effects in context by considering some lessons from the recent fiscal literature and other literature that uses local data to answer macro questions.

The first lesson from the literature is that local effects, which the authors estimate, are not the same as aggregate effects (Nakamura and Steinsson 2014). Because the panel data estimation uses year-fixed effects, all aggregate effects are removed. Thus, the estimates of employment effects are *relative* effects that answer the following question: If commuting zone A receives $1 more than the other commuting zones, how much does employment change in commuting zone A relative to the average? This estimate is different from the aggregate effect in several ways. First, it does not incorporate the added general equilibrium effects due to tax or deficit financing. Second, it may be smaller or larger than the aggregate estimate: it may be smaller if there are positive spillovers across commuting zones, but it may be larger if there are "business stealing" effects across commuting zones.

A second insight that comes out of the infrastructure literature is the implication of time-to-spend and time-to-build delays that are inherent to infrastructure projects. Figure 1 shows the cumulative percentage of total ARRA infrastructure appropriations that were spent each year. Despite the emphasis on "shovel-ready projects," only 10 percent of appropriations were spent by summer 2009 and under 60 percent had been spent by the following summer. As I now demonstrate, if there were also delays in green infrastructure spending, then the minimal short-run effects estimated by the authors can be explained.

As Leeper, Walker, and Yang (2010) and others have argued, these delays severely reduce the short-run stimulus effects of infrastructure investment. In Ramey (2021), I show that even in a medium-scale New Keynesian model, calibrated to give high multipliers for government consumption spending, infrastructure spending with time-to-spend and time-to-build delays offers little short-run stimulus. To illustrate the effects, figure 2 shows the simulations from a version of the New Keynesian model used in my work. The graphs show the effects of either a shock to government

Figure 1. Federal Highway Administration Outlays from the ARRA: Cumulative Percentage Spent of Total Appropriation

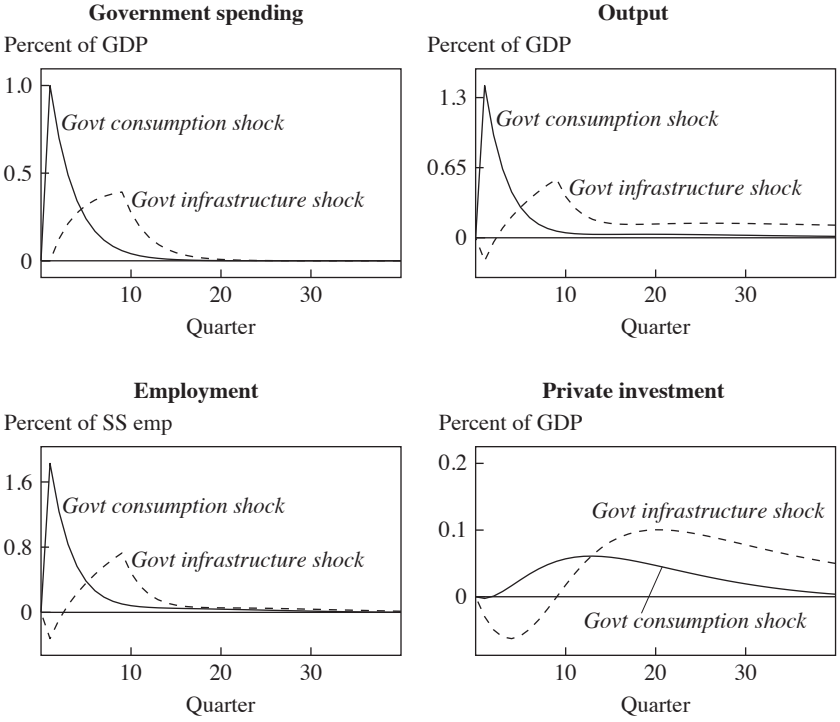


Source: Leduc and Wilson (2017).
Note: State-level data from replication files are aggregated to the national level.

consumption or government infrastructure investment appropriations. In both cases, the appropriations process is assumed to be serially correlated so that the initial bump to appropriations is followed by smaller additions to appropriations until they return to zero after about three years. In the case of government consumption spending, appropriations translate immediately into government spending. However, in the case of government infrastructure spending, there are both time-to-spend and time-to-build delays. In particular, a one-unit rise in appropriations is spread out equally over the next eight quarters. In addition, infrastructure does not become productive until eight quarters after the beginning of construction.

Figure 2 shows the paths of government spending, output, employment, and private investment for each of the two experiments. These graphs show that a shock to government consumption appropriations equal to 1 percent of GDP leads output to rise by more than 1.3 percent, implying an impact multiplier above 1.3. Employment rises 1.7 percent relative to its steady state. Private investment builds up slowly because of the assumption of adjustment costs to investment. Overall, the rise in government consumption has significant short-run stimulus effects in this New Keynesian model. As discussed in Ramey (2021), the three most important features of the model for the short-run stimulus effects are adjustment costs on private

Figure 2. Macroeconomic Effects of Government Consumption Spending versus Infrastructure Spending with Delays in a New Keynesian Model



Source: Ramey (2021).

Note: These graphs show impulse responses to a shock to government appropriations destined to either government consumption or government infrastructure investment. The impulse responses are based on simulations of the medium-scale New Keynesian model.

investment (which prevent the standard crowding out), the presence of "hand-to-mouth" households which always consume 100 percent of their income, and that employment is mostly determined by labor demand and not labor supply.

The effects are very different for infrastructure spending because of the time-to-spend and time-to-build lags. Because of the spending delays, output and employment do not rise on impact and in fact fall slightly. Without the government spending, firms do not demand more labor, so the hand-to-mouth households do not experience rises in income and hence do not consume more. Once spending ramps up, there is a slow rise in output and employment. Private investment falls in the short run and then rises more in the intermediate run. The first lesson from these simulations is that

15985-01b-Popp_Comm&GD-4thPgs.indd 60 5/12/22 12:10 PM

time-to-spend and time-to-build delays prevent government infrastructure spending from acting as a short-run stimulus.

The second lesson from the simulations is the theoretical prediction about the long-run effects on employment. As figure 2 shows, output is predicted to remain above its steady state out through ten years (forty quarters) because of the higher level of public capital. However, employment returns to its steady state after four years. Both a standard neoclassical model and a New Keynesian model predict that a relatively transitory rise in either government consumption or government investment should have no noticeable effect on long-run employment. This result should be kept in mind when considering the plausibility of the authors' estimates suggesting more permanent employment effects.

SUMMARY OF THE AUTHORS' ESTIMATES WITH A COMPARISON TO THE EFFECTS OF THE ARRA HIGHWAY CONSTRUCTION SPENDING Popp and colleagues show (in their figure 2) the estimates of the relative employment effects of the green spending in the ARRA based on their commuting zone data. Their estimates imply about ten jobs created per $1 million in the first year of the ARRA in 2009, gradually rising to a plateau of about twenty-five jobs per $1 million by 2017. They conclude that the ARRA green spending, which was temporary, may have led to permanent job creation.

This pattern of slowly rising effects of jobs created at longer horizon contrasts with results for highway spending. For example, Garin (2019) studies the job creation effects of the road and highway spending part of the ARRA at the county level. Focusing first on construction employment, he finds that the spending raises highway construction employment little the first year (2009), creates a peak of two jobs per $1 million in 2010, and then slowly tapers back to zero. When he studies total payrolls (in dollars), he finds no effect the first several years, then a rising effect that peaks in 2013 before declining again. Thus, the authors' finding of an ever-rising job creation rate contrasts with Garin's (2019) results for the highway construction parts of the ARRA. One possible source of the permanent estimates in the authors' analysis is the presence of pre-trends. As they highlight, and is evident in their figure 2, it appears that there are significant pre-trends. Thus, the estimates of the long-run effects could reflect that the commuting zones that received more green energy spending had a higher trend rate of employment growth irrespective of the ARRA green spending.

In contrast, the authors' estimates of employment effects by category are on firmer ground, since there are no significant pre-trends. Employment in manual occupations, green jobs, and construction jobs share a similar pattern: little or no effect for the first three years after the initial spending

but then rising to a higher plateau from around 2012 through 2017. These higher intermediate-run effects are quite different from the more transitory effects Garin (2019) finds for highway spending.

These more persistent effects on employment categories at the commuting zone level found by the authors are not necessarily at odds with the dynamic general equilibrium analysis shown above. The dynamic general equilibrium analysis shows that the effects on *aggregate* employment should last only during the transition. That does not mean, however, that there cannot be permanent reallocations of employment across job categories or geographic levels. The next section, however, suggests that an important additional green program may be confounding the effects.

SOLAR TAX INCENTIVES: POSSIBLE CONFOUNDING EFFECTS AND JOB CREATION POTENTIAL The authors focus on the green spending elements of the ARRA but do not consider tax incentives for green energy since such are beyond the scope of the paper. I believe that these additional incentives cannot be ignored because they are likely to confound the estimates of the effects of the spending program.

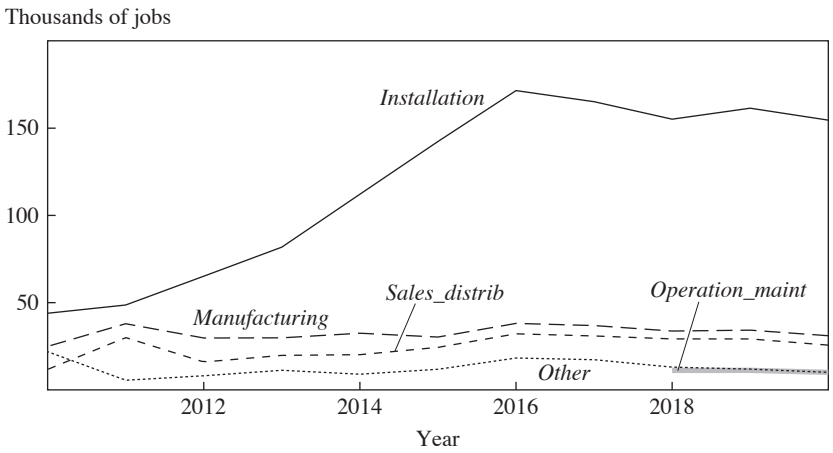
Residential and commercial investment federal tax credits for rooftop solar were adopted in the mid-2000s and subsequent legislation has extended them multiple times. The tax credit was 30 percent of the installation cost of rooftop solar until recently, when it was reduced to 26 percent.[1] Thus, the tax incentive regime completely overlaps the period of analysis by the authors. Moreover, the price of photovoltaic cells fell 70 percent in the last decade, leading to an upward trend in the incentive to take advantage of the rooftop solar tax incentives (Solar Energy Industries Association 2021).

Figure 3 shows the growth of solar jobs by category from 2010 to 2020 from the National Solar Jobs Census (Solar Energy Industries Association 2021). The graph shows that solar installation jobs are the dominant source of both the level and growth of total solar jobs. Installation jobs currently account for two-thirds of all solar jobs. Important to note is the time pattern: job growth was slow initially and then took off after 2012, reaching a plateau starting around 2016. This pattern is very similar to the pattern of estimates by the authors for the effects of the ARRA green spending. They interpret their results as indicating that the ARRA appropriations led to

1. The history through 2016 is summarized on the Solar Energy Industries Association website, https://www.seia.org/sites/default/files/resources/History%20of%20ITC%20Slides.pdf, and the more recent extensions are discussed at Office of Energy Efficiency and Renewable Energy, "Homeowner's Guide to the Federal Tax Credit for Solar Photovoltaics," https://www.energy.gov/eere/solar/homeowners-guide-federal-tax-credit-solar-photovoltaics and https://news.energysage.com/solar-tax-credit-explained/.

15985-01b-Popp_Comm&GD-4thPgs.indd 62 5/12/22 12:10 PM

Figure 3. Solar Employment by Sector



Source: Solar Energy Industries Association (2021).

permanent job creation. I suggest an alternative hypothesis: the commuting zones that received ARRA green spending appropriations were also areas that were ideal for solar. As a result, households and businesses in those areas were more likely to take advantage of the solar tax incentives, which remained in place long after the ARRA green spending was spent. Thus, the apparent permanent job creation effects of the ARRA green spending are more likely to have been due to the tax incentives for rooftop solar interacting with the declining price of photovoltaic cells.

It would be interesting for future research to study the effects of rooftop solar for an additional reason: its use of labor and land. Installing rooftop solar requires more labor hours per energy unit installed. For example, according to the 2015 National Solar Jobs Census, residential installations required 40 labor hours per 5 kilowatt hours, commercial installations required 36, and utility-scale installations 25 (Solar Foundation 2016, 24). Many have argued that utility-scale installations, which involve large solar farms often situated in remote desert areas, are superior because they take advantage of economies of scale. However, this argument neglects two factors. First, rooftop solar reduces the need for transmission lines to carry the energy across long distances. Second, rooftop solar uses land more efficiently and does not disturb endangered habitat. A recent Brookings study highlights the fact that wind and solar energy require at least ten times more land per unit of power produced than fossil fuels do (Gross 2020). This

demand for additional land leads to numerous conflicts over land use. For example, in California climate change activists have clashed with conservationists over the construction of large solar farms in areas with endangered species (Roth 2021). In contrast, rooftop solar is installed on existing buildings, near the ultimate user.

THE IMPORTANCE OF COMMUNITY-PRESERVING POLICIES As I mentioned in the introduction to my comments, the major motivation for the authors' analysis is determining whether the workers displaced by the decline of fossil fuels can be reemployed in green industries. I have already raised questions about their finding of seemingly permanent job gains from temporary green spending. If the spending does not lead to permanent job gains, then there can be serious impacts on communities.

Consider the example of the mining communities of Asturias, Spain. The Spanish government phased out mining in the Asturias region of Spain, which had been mining coal for hundreds of years, and promised to bring in green jobs and retraining programs. According to numerous reports, the government has not fulfilled its promise. Instead, it started importing cheap coal from China. The older coal miners were pensioned off, preventing large income losses. However, because there are no jobs for the younger people, the towns in the mining areas are being depopulated, with mostly older people left behind (Benavides 2019).

The lesson to be learned from the example of Asturias, Spain, is that without the prospect of permanent, good-paying jobs for younger people, communities decline. The cycle can feed on itself since depopulation leads to reduced local tax revenue. The green transition is likely to face formidable opposition from some quarters if governments cannot develop credible, community-preserving policies.

CONCLUSIONS Popp and colleagues have written a very useful analysis of the possibilities of the green transition providing jobs to fossil fuel and other workers. In my discussion, I have raised questions primarily about the estimated permanent effects of the temporary green spending stimulus in the ARRA. Theory predicts that there should be no permanent employment effects of temporary spending at the aggregate level, but since the authors' estimates are at the local level, permanent employment shifts are possible. However, I have also raised the possibility that their estimates may be picking up the effects of another government green incentive, tax credits for rooftop solar. That incentive has been in place since the mid-2000s and, in conjunction with the declining price of photovoltaic panels, has likely led to an upward trend in green employment. If the ARRA green spending was more likely to be directed to communities that also had more

15985-01b-Popp_Comm&GD-4thPgs.indd 64 5/12/22 12:10 PM

natural solar potential, then the upward trends in job creation found by the authors may be picking up these alternative effects.

REFERENCES FOR THE RAMEY COMMENT

Benavides, Lucía. 2019. "Spain's Coal Miners Continue to Wait for Their Country's 'Green New Deal.'" *The World*, August 16. https://theworld.org/stories/2019-08-16/spains-coal-miners-continue-wait-their-countrys-green-new-deal.

Garin, Andrew. 2019. "Putting America to Work, Where? Evidence on the Effectiveness of Infrastructure Construction as a Locally Targeted Employment Policy." *Journal of Urban Economics* 111:108–31.

Gross, Samantha. 2020. *Renewables, Land Use, and Local Opposition in the United States*. Washington: Brookings Institution, Foreign Policy. https://www.brookings.edu/research/renewables-land-use-and-local-opposition-in-the-united-states/.

Leduc, Sylvain, and Daniel J. Wilson. 2017. "Are State Governments Roadblocks to Federal Stimulus? Evidence on the Flypaper Effect of Highway Grants in the 2009 Recovery Act." *American Economic Journal: Economic Policy* 9, no. 2: 253–92.

Leeper, Eric M., Todd B. Walker, and Shu-Chu Susan Yang. 2010. "Government Investment and Fiscal Stimulus." *Journal of Monetary Economics* 57, no. 8: 1000–1012.

Nakamura, Emi, and Jón Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review* 104, no. 3: 753–92.

Ramey, Valerie A. 2021. "The Macroeconomic Consequences of Infrastructure Spending." In *Economic Analysis and Infrastructure Investment*, edited by Edward L. Glaeser and James M. Poterba. Chicago: University of Chicago Press.

Roth, Sammy. 2021. "Two Biden Priorities, Climate and Conservation, Collide in the California Desert." *San Diego Union Tribune*, January 21.

Solar Energy Industries Association. 2021. *National Solar Jobs Census 2020*. Washington: Solar Energy Industries Association. https://www.seia.org/research-resources/national-solar-jobs-census-2020.

Solar Foundation. 2016. *National Solar Jobs Census 2015*. Washington: Interstate Renewable Energy Council. https://irecusa.org/resources/national-solar-jobs-census-2015/.

GENERAL DISCUSSION Jason Furman asked about the implications of the findings of the paper for welfare analysis. He noted that, generally, in educational settings students are taught to ignore jobs in cost-benefit analyses or even to treat jobs as a cost, given that the presence of jobs indicates a need to divert resources to that program (though this treatment depends on whether or not the economy is in a recession). He additionally asked if the implications for welfare or policy analysis change depending on whether the economy is in a recession or is closer to full employment.

Laurence Meyer critiqued the use of the American Recovery and Reinvestment Act (ARRA) as the framework for the paper, as the law was a deficit-financed stimulus package, and future green packages would likely have to be at least partially paid for through tax increases. He also suggested that the key benefits of green fiscal spending do not come from employment gains but rather from addressing the climate crisis.

David Popp addressed the questions from Furman and Meyer, stating that the central purpose of the paper was to respond to existing policy proposals to address climate change and to discuss the political feasibility of these proposals. While employment may not be as important a goal as emissions reductions when evaluating the effectiveness of green spending, it does impact the political feasibility of passing climate policy, he concluded.

Robert Hall suggested that the ARRA's effects on employment would comprise only a small portion of total effects due to the low elasticity of labor supply. He posited that the effects on wages may be stronger than the effects on employment and encouraged the authors to explore the impact of green spending on wages.

Olivier Blanchard noted that the nature of labor and skills required by green jobs may change over time, especially since many green projects are temporary. As an example, he remarked that proposals to expand the electric vehicle charging infrastructure in the United States would likely generate strong demand for local labor but that this demand would fade after charging stations have been installed. He expressed concern over applying the results from the paper to predict the impact of future policies, as doing so would require knowledge of the skills needed for future green jobs.

Heather Boushey echoed Blanchard's concerns that applying lessons from this paper to future policies requires the ability to identify green jobs in the future. For instance, she noted that all jobs in auto manufacturing could be categorized as a green job in the future, since many expect that all new cars sold will be zero-emission vehicles.[1] She also revisited Meyer's concern that the lessons learned from analyzing the employment impact of

1. In California, the governor issued an executive order; Office of Governor Gavin Newsom, "Governor Newsom Announced California Will Phase Out Gasoline-Powered Cars and Drastically Reduce Demand for Fossil Fuel in California's Fight against Climate Change," September 23, 2020; https://www.gov.ca.gov/2020/09/23/governor-newsom-announces-california-will-phase-out-gasoline-powered-cars-drastically-reduce-demand-for-fossil-fuel-in-californias-fight-against-climate-change/. See also Neal E. Boudette and Coral Davenport, "G.M. Will Sell Only Zero-Emission Vehicles by 2035," *New York Times*, January 28, 2021, https://www.nytimes.com/2021/01/28/business/gm-zero-emission-vehicles.html; and Honda, "Summary of Honda Global CEO Inaugural Press Conference," April 23, 2021, https://global.honda/newsroom/news/2021/c210423eng.html.

15985-01b-Popp_Comm&GD-4thPgs.indd 66 5/12/22 12:10 PM

green spending from the ARRA may not apply to current and future green spending thanks to two key distinctions between the two policies. First, the ARRA occurred in a weaker economy, and green spending will occur in an economy that is closer to full employment. Second, the ARRA was focused on stimulus, while green spending will likely center on fostering innovation and effecting structural transformation.

Addressing Boushey and Blanchard, Popp agreed that readers should be careful when generalizing lessons from 2009 to understand current policies and policy proposals. For example, as pointed out in the paper, much of the focus on energy retrofitting in the ARRA was to bolster the construction industry. He argued that the implications from this paper center on the importance of being aware of skills changing and determining which skills line up with policy needs at any given moment.

Popp responded to Hall's question regarding the ARRA's impact on wages and said that the authors had already estimated the impact of ARRA green spending on wages but found no relationship in the short run or long run. These results had been excluded from the presentation due to time constraints.

Michael Kiley commented that the paper was framed to look at the effect of stimulus on employment. However, he commented that most present and future investments in green technologies—including President Biden's Build Back Better agenda—will not necessarily increase the level of capital as the proposals are instead focused on reallocating activity or changing the nature of capital investment toward, for example, greener activities, rather than focused on increasing the level of activity. He wondered if the authors could shift the focus of the paper to discuss how to make reallocation most effective.

Caroline Hoxby noted that the classification of green jobs versus non-green jobs can be arbitrary. While the Bureau of Labor Statistics has a definition of green jobs, considerable judgment is necessary to apply it.[2] Only a portion of jobs are *obviously* green—for instance, a wildlife conservation or stormwater management worker. Others are harder to categorize—such as a worker who enforces environmental regulations among numerous other regulations. The more incentives exist to report a job as green, the more that firms will find ways to reclassify jobs as green. Therefore, a comprehensive study of the impact of green spending may have to examine both categories as well as employment overall.

2. See US Bureau of Labor Statistics, "Green Jobs Overview," https://www.bls.gov/green/overview.htm.

James Stock argued that the authors should control for renewable portfolio standards, as renewable portfolio standards were one of the two most active policies used to address climate change (alongside ARRA subsidies) at the time.[3] However, he mentioned that this control might be more important in theory but not as impactful in practice.

Popp first responded to Stock's question on renewable portfolio standards. He argued that since all renewable portfolio standards began before the 2008 recession, the paper's fixed effects should control for the impact of those policies. He noted that the authors also controlled for wind potential, which would also relate to renewable portfolio standards. However, he also observed that certain smaller policies for renewable energy, such as solar tax credits which had been mentioned by discussant Valerie Ramey, may be harder to control for.

Popp also acknowledged the validity of the comment from Hoxby on defining the green jobs of the future. He stated that responding to that issue might be best accomplished through future research rather than through this paper.

Citing a paper by Sanya Carley, Sean Nicholson-Crotty, and Eric J. Fisher, Popp addressed the comment by discussant Gabriel Chodorow-Reich and noted that there is often a gap between annual allocations and realized expenditures due to bureaucratic requirements.[4] For example, he mentioned that "Made in America" requirements hampered energy retrofitting projects because it was difficult to find LED light bulbs that had been manufactured in America. Thus, both the effects of allocated money or of actual spending on a yearly basis could be valid measures. One measures what was intended and the other measures how the program was actually implemented.

Francesco Vona added that an important contribution of the paper was to extend a measurement framework that explores the skills requirements for green jobs of the future, based on the previous paper by Vona and others (2018).[5] This framework is rooted in the task-based framework to labor

3. R. Wiser, G. Barbose, J. Heeter, T. Mai, L. Bird, M. Bolinger, and others, *A Retrospective Analysis of the Benefits and Impacts of U.S. Renewable Portfolio Standards* (Washington: National Renewable Energy Laboratory and Lawrence Berkeley National Laboratory, 2016), http://www.nrel.gov/docs/fy16osti/65005.pdf.

4. Sanya Carley, Sean Nicholson-Crotty, and Eric J. Fisher, "Capacity, Guidance and the Implementation of the American Recovery and Reinvestment Act," *Public Administration Review* 75, no. 1 (2015): 113–25, doi: 10.1111/puar.12294.

5. Francesco Vona, Giovanni Marin, Davide Consoli, and David Popp, "Environmental Regulation and Green Skills: An Empirical Exploration," *Journal of the Association of Environmental and Resource Economists* 5, no. 4 (2018): 713–53.

15985-01b-Popp_Comm&GD-4thPgs.indd 68 5/12/22 12:10 PM

markets that has been advanced by David Autor and others.[6] He added that an additional contribution was a structure for measuring reallocation costs based on "skill distances." Finally, he argued that the basic framework they employ will be useful in calibrating models of the effects of climate change with friction, where reallocation is frequent but employment effects are small overall.

6. David H. Autor, "The 'Task Approach' to Labor Markets: An Overview," working paper 18711 (Cambridge, Mass.: National Bureau of Economic Research, 2013), http://www.nber.org/papers/w18711; David H. Autor, Frank Levy, and Richard J. Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics* 118, no. 4 (2003): 1279–333.

15985-01b-Popp_Comm&GD-4thPgs.indd 70 5/12/22 12:10 PM

SHELBY R. BUCKMAN
Stanford University

MARY C. DALY
Federal Reserve Bank of San Francisco

LAURA Y. CHOI
Federal Reserve Bank of San Francisco

LILY M. SEITELMAN
Boston University

The Economic Gains from Equity

ABSTRACT How much larger would the US economic pie be if labor market outcomes were more equitably distributed by race and ethnicity? Using data from the Current Population Survey (1990–2019), we estimate the improvements in labor contribution to aggregate output associated with making the outcomes for Black, Hispanic, and other minority groups at least as favorable as those for non-Hispanic white individuals in employment, hours worked, educational attainment, educational utilization, and earnings. We find significant economic gains, measured in trillions of dollars of GDP. Our results indicate that ensuring all Americans have an equitable opportunity to participate in the economy is an economically significant way to increase aggregate prosperity.

In 1964 President Lyndon B. Johnson signed the Civil Rights Act, prohibiting discrimination based on race, color, religion, sex, or national origin. The signing marked a shift in American law that was intended to

Note: This paper was guest-edited by Katharine Abraham.

Conflict of Interest Disclosure: Shelby R. Buckman is a graduate student in economics at Stanford University and a past employee at the Federal Reserve Bank of San Francisco, during which time she did the majority of her contribution to this paper. Laura Y. Choi is senior vice president of public engagement at the Federal Reserve Bank of San Francisco. Mary C. Daly is president and chief executive officer of the Federal Reserve Bank of San Francisco, CRIW member, and IZA research fellow. Lily M. Seitelman is a graduate student in economics at Boston University and a past employee at the Federal Reserve Bank of San Francisco, during which time she did the majority of her contribution to this paper. 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 Federal Reserve Bank of San Francisco. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under grant no. DGE-1840990 and grant no. DGE-1840990.

15985-02a_Buckman-6thPgs.indd 71 5/12/22 12:10 PM

remove barriers and deliver more equal opportunities, including in the labor market. Today, significant gaps in outcomes by race and ethnicity remain. Indeed, an extensive literature documents persistent, and even widening, disparities in employment and earnings in the United States that leave Black, Hispanic, and other minorities well behind their non-Hispanic white counterparts. The research also shows these differences in outcomes, the gaps, are only partially explained by measurable differences in education, experience, and other characteristics related to productivity.[1] The result is that nearly sixty years after the passage of the Civil Rights Act, and despite the countless policies and programs that have followed, race and ethnicity remain significant predictors of labor market success in the United States (Rodgers 2019).

We argue that these facts present an economic problem.[2] If talent and innate preferences are evenly distributed by race and ethnicity—a premise that seems hard to refute—persistent racial and ethnic disparities represent misallocation and lost production.[3] In other words, the persistence of systemic disparities is costly, and eliminating them has the potential to produce large economic gains (Hsieh and others 2019; Turner 2018; Treuhaft, Scoggins, and Tran 2014).

Of course, the benefits go well beyond the inclusion of previously sidelined or underutilized resources. The opportunity to use one's talents fully, unbridled by prejudgment or other artificial barriers, is at the foundation of a dynamic economy. Individuals invest in themselves based on the returns they expect to receive. Persistent barriers and systemically lower payoffs to effort depress these incentives, potentially leading to lower human capital investment and further solidification of gaps in outcomes.[4] In other words, more equitable outcomes matter both for the level of GDP and for the process of sustained economic growth. This means that changing opportunity affects both current and future economic output.

1. This literature spans multiple disciplines. For example, see Williams and Wilson (2019), Wilson and Jones (2018), Cajner and others (2017), Daly, Hobijn, and Pedtke (2017), and Pager and Shepherd (2008).

2. Several researchers, including Williams and Spriggs (1999), Mason and Williams (1997), and Darity and Williams (1985), have noted that standard economics models would not naturally reach this conclusion. Our paper builds on their insights. We elaborate on their points later in the discussion.

3. We are not the first to recognize this; see Bostic (2020), Cook (2020), Peterson and Mann (2020), and Daly (2021).

4. Although not the subject of this paper, these barriers also lead to gaps by race and ethnicity in consumption, savings, and wealth, which leave individuals, families, and communities more vulnerable to economic shocks; see Bhutta and others (2020).

15985-02a_Buckman-6thPgs.indd 72 5/12/22 12:10 PM

In this paper, we offer some initial estimates of what the gains to the labor contribution to GDP might be from equalizing labor market outcomes across race and ethnicity.[5] Specifically, using data from the Current Population Survey (CPS, 1990–2019), we estimate the impact on aggregate output of making the labor market outcomes of Black, Hispanic, and other minority groups at least as favorable as those of the non-Hispanic white population, the long-standing majority group in the United States.[6] Our findings point to considerable gains, measured in trillions of dollars of GDP. We then ask whether the gains from more equal outcomes have changed over time. We find that the benefits from equalizing outcomes have risen, owing to the persistence of economic disparities and the rising share of the population that experiences them. Finally, we consider the extent to which closing each outcome gap would contribute to increases in overall output. Out of the measures we consider, we find that eliminating racial and ethnic disparities in employment rates and educational attainment makes the largest contributions. Additional meaningful gains could come from eliminating residual earnings gaps not explained by these and other productivity-related indicators.

The remainder of the paper is structured as follows. We begin by documenting gaps in measures of labor market success by race and ethnicity over the past three decades in section I. We then briefly discuss the related literature, calling out where our work contributes in section II. In section III, we lay out our framework and describe our results. We conclude in section IV with a discussion of possible further considerations and research, including issues policymakers, the private sector, and American society will have to grapple with to obtain the kinds of gains we document.

I. Labor Market Outcomes by Race and Ethnicity

Although gaps in labor market outcomes by race and ethnicity are often discussed and fairly well known, the extent of the gaps and their persistence bears reviewing. To do this we use CPS data from 1990 through 2019.[7] To avoid concerns about differences in schooling or retirement behavior, we focus on civilian noninstitutionalized adults age 25–64. We further

5. Examples of similar exercises can be found in Peterson and Mann (2020), Noel and others (2019), and Turner (2018).

6. We consider this a reasonable benchmark to make the point that unequal labor market outcomes translate into lower aggregate output and that greater equity increases overall prosperity.

7. We restrict our analysis to the outgoing rotation groups of the CPS basic monthly files.

restrict our baseline sample to people who are not self-employed.[8] We divide this population into eight mutually exclusive groups defined by gender and race or ethnicity.[9] Although there is no perfect way to categorize individuals by race or ethnicity, we follow what is commonly done in the literature. Specifically, we define four mutually exclusive race/ethnicity groups based on self-reported designations: non-Hispanic white, non-Hispanic Black, Hispanic, and all remaining non-Hispanic, non-white, and non-Black individuals (e.g., Asian, Pacific Islander, American Indian).[10] Throughout the paper we refer to these groups as white, Black, Hispanic, and API+, respectively.

To assess relative labor market success across groups, we consider five metrics: employment, hours worked, earnings, educational attainment, and educational utilization. We define employment using the variable *monthly labor force recode*, restricting our sample to the values "employed-at work" and "employed-absent." We aggregate these responses annually and compute the average share of employed individuals in each year. Conditional on being employed, we then compute hours worked and earnings. We define hours worked as usual weekly hours worked over all jobs, and we similarly average this value over our sample and on an annual basis. We measure earnings as average hourly earnings, defined as usual weekly earnings divided by usual weekly hours adjusted for inflation using the personal consumption expenditures price index.[11]

We categorize educational attainment as the maximum education completed across four mutually exclusive categories: high school or less, some college, bachelor's degree, and postgraduate studies. Finally, we compute a measure of educational utilization for employed individuals with a bachelor's degree or higher. This measure is intended to capture the fact

8. We make this restriction due to a change in the collection of hours data in 1994 for the self-employed that limits the accuracy of comparisons over time. In analysis not shown, we repeat everything to include the self-employed along with all other employed people and the results are qualitatively similar.

9. The question in the CPS files that we use to define our gender variable asks, "What is _____'s sex?" In keeping with the literature, we refer to the distinctions of men and women as gender.

10. We recognize that the final group is quite heterogeneous. Unfortunately, given the limits of sample size in the CPS data, we are unable to reliably divide this group into smaller and more representative categories.

11. To account for discontinuities in CPS top-coding, we adjust top-coded earnings following methods developed by the Center for Economic Policy Research (CEPR Data, CPS ORG Programs, https://ceprdata.org/cps-uniform-data-extracts/cps-outgoing-rotation-group/cps-org-programs/). Employed individuals with missing earnings values are assigned zero earnings.

15985-02a_Buckman-6thPgs.indd 74 5/12/22 12:10 PM

that many racial and ethnic minorities are overeducated for the jobs they hold (Williams and Wilson 2019; Rose 2017).[12] We measure educational utilization for individuals by comparing their educational attainment to what is required by their occupation. For simplicity we define required education as needing or not needing a bachelor's degree. Following Williams and Wilson (2019), we estimate required education, at the four-digit occupation level, by computing the percentage of individuals in that occupation that have less than a bachelor's degree (e.g., with some college or high school or less). If 50 percent (or more) of workers in a four-digit occupation group have less than a bachelor's degree, we classify that occupation as not requiring a bachelor's degree. We repeat this for each year in our sample. We then compare these requirements to the educational attainment of the workers in our sample. Workers with more education than their occupation requires are considered underutilized. All other workers are classified as fully utilized.

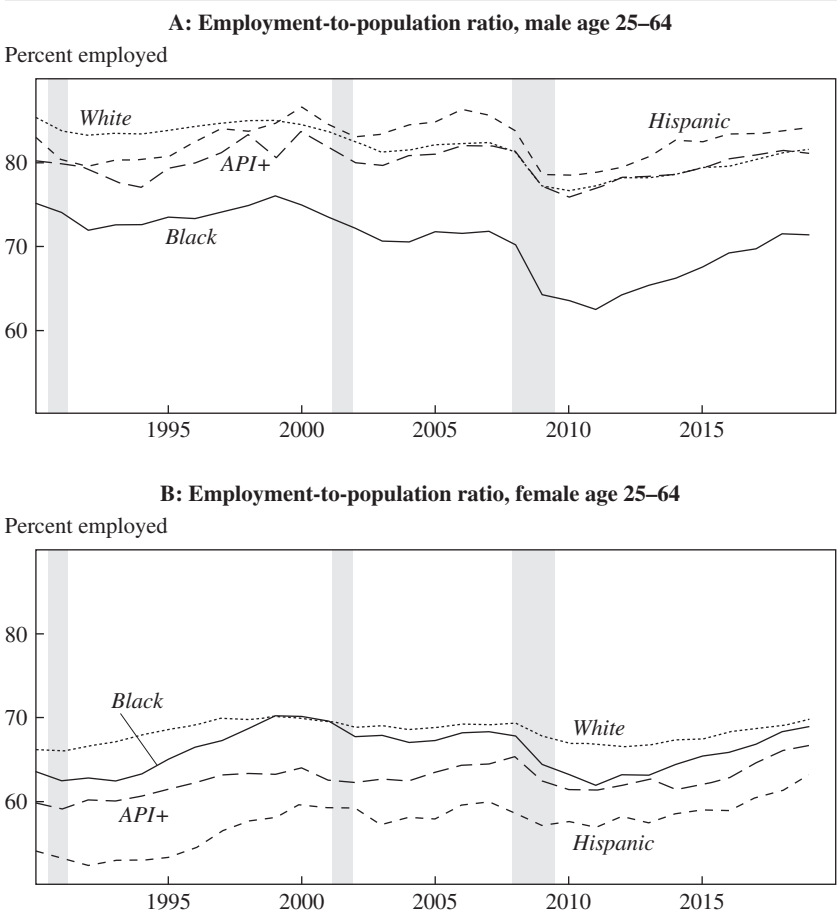
I.A. Levels and Trends

We begin with employment, shown in figure 1, which plots the percentage of the population age 25–64 employed in each year by gender and race. Several things stand out in these charts. First, employment rates for all fluctuate with the business cycle, moving up in good times and down in bad times. Second, there are considerable differences by gender. Among men, shown in panel A, the secular trend in employment is downward, meaning that employment rates are lower today than they were in 1990. Looking more specifically at differences by race and ethnicity, clear differences emerge. Employment rates for Black men are consistently lower than for all other groups. For example, in 2019 the employment rate of white, Hispanic, and API+ men was just over 80 percent; the employment rate for Black men was just over 70 percent. Over the entire sample period, the average employment gap between white and Black men is about 11 percentage points. Although the gap between Black and white men narrows somewhat during expansions (Aaronson and others 2019), the main finding is that Black men are far less likely to be employed than other men.

The pattern for women is quite different. Overall, women have lower employment rates than their male counterparts. Among women, white women have higher employment rates than their Black, Hispanic, or API+

12. Related literature finds that barriers to entry for racial and ethnic minorities have resulted in a misallocation of talent across industry and occupation (Hsieh and others 2019; Bell and others 2019).

Figure 1. Trends in Employment

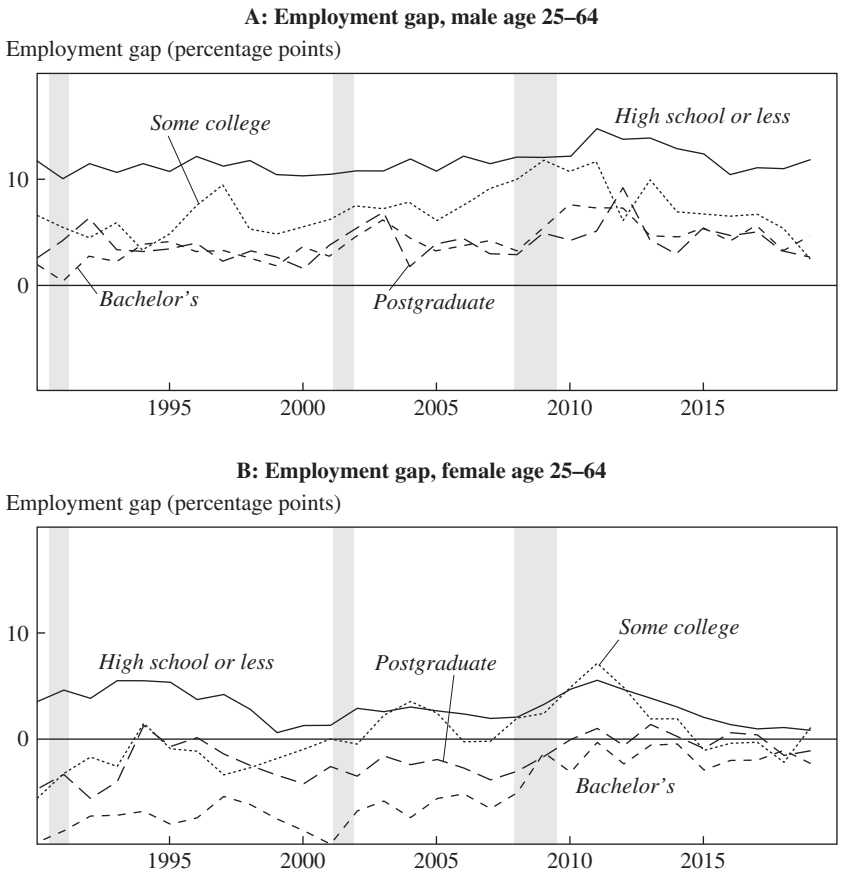


Source: Authors' calculations using CPS data.

counterparts. Black women have the next highest employment rates, although the gap between white and Black women fluctuates considerably over the business cycle, widening in recessions and narrowing again well into expansions. Employment rates since the Great Recession for API+ women are similar to those of Black women, however they exhibit very little cyclicality over the sample. Hispanic women's employment rates are substantially lower and have been hovering slightly below 60 percent since 2000. The exception to this is in the final years of the last expansion. Between 2016 and 2019, employment rates for Hispanic women rose 4.3 percentage points.

15985-02a_Buckman-6thPgs.indd 76 5/12/22 12:10 PM

Figure 2. White/Black Trends in Employment Gaps by Education



Source: Authors' calculations using CPS data.

Importantly, as many have noted, these differences in employment do not fully owe to differences in education (Spriggs and Williams 2000; Williams and Wilson 2019; Daly, Hobijn, and Pedtke 2017; Cajner and others 2017). This can be seen in figures 2, 3, and 4, which show how employment rates across groups differ regardless of education level. The charts plot employment rate gaps by education between white men and women and their Black, Hispanic, and API+ counterparts, for example, the percentage point difference between white male employment rates and Black male employment rates.

Starting with figure 2, it is clear that educational attainment alone does not close gaps in employment between white and Black men. That said,

gaps do shrink with increased educational attainment. For example, in 2019 the employment gap for Black men with a bachelor's degree was about 5 percentage points, compared to over 11 percentage points for Black men with high school or less. Although the gaps have fluctuated somewhat over time, by and large they are the same in 2019 as they were in 1990. This holds for all education levels. As was the case in the more aggregated trends, the pattern for women is different. For much of the sample, Black women with a bachelor's degree or postgraduate education have had higher employment rates than similarly educated white women. The gaps in employment have been falling over time, although Black women with a bachelor's degree remain slightly more likely to be employed. In contrast, Black women with some college or high school or less education have frequently had lower employment rates than their white counterparts. These gaps have also closed over time.

Figure 3 shows the same plot for Hispanic men and women. Recall that Hispanic men generally have higher employment rates than white men. The figure shows that this is especially true for Hispanic men with high school or less. Hispanic men with a bachelor's or postgraduate degree are employed at about the same rate as their white counterparts, resulting in a minimal employment gap. Similar to the pattern for Black women, the employment gaps for Hispanic women at all education levels have been converging toward zero, although some gaps remain.

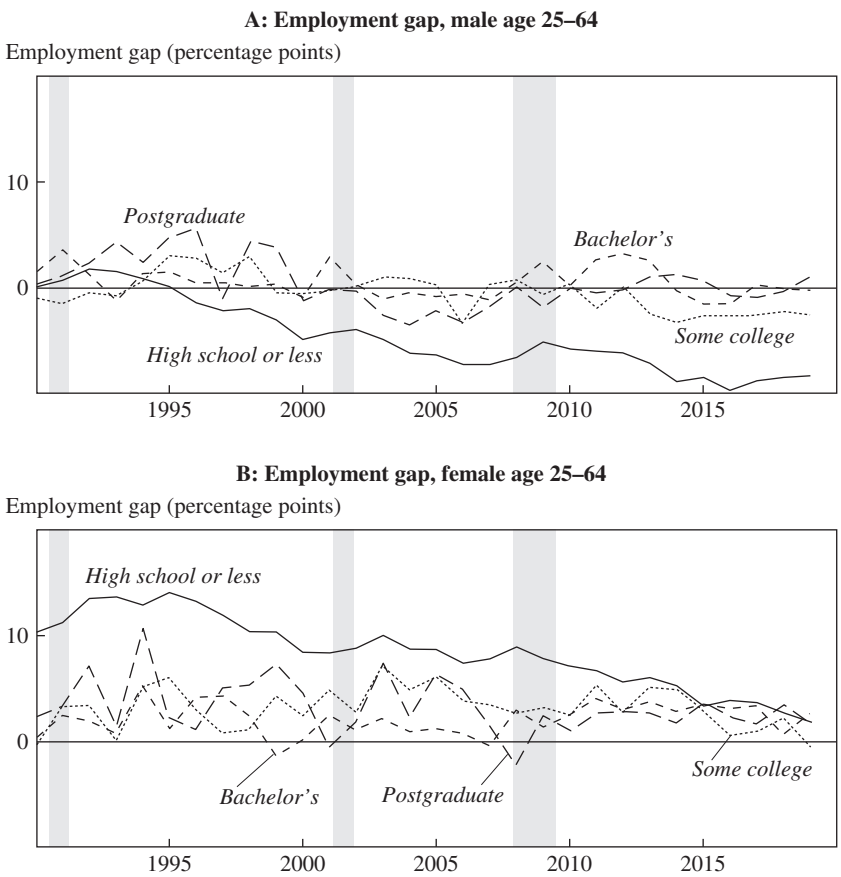
The data in figure 4 show employment gaps for API+ men and women. For API+ men, the gaps have been trending down over time for all education levels. However, gaps do remain for API+ men with less than a postgraduate degree.[13] The pattern by education is completely reversed for more educated API+ women, with larger employment gaps and lower employment rates.

Turning to hours (figure 5), among those who are employed, white men work more hours than Black, Hispanic, or API+ men. In the early 2000s, the gap between white and API+ men began to decrease, while the values for both groups mostly leveled off in the recent expansion. The gaps in hours between white men and Black and Hispanic men have begun to narrow, as hours for Black and Hispanic men have risen. The growth in hours worked for these two groups was especially noticeable in the last expansion. In 2019, the hours gap was at the lowest level recorded over the period, at about 1.7 hours per week between Black and white men, and 1.4 hours per week between Hispanic and white men. Again, for women, the patterns

13. As mentioned earlier, the majority of the API+ sample is Asian American, with higher than average levels of educational attainment.

15985-02a_Buckman-6thPgs.indd 78 5/12/22 12:10 PM

Figure 3. White/Hispanic Trends in Employment Gaps by Education

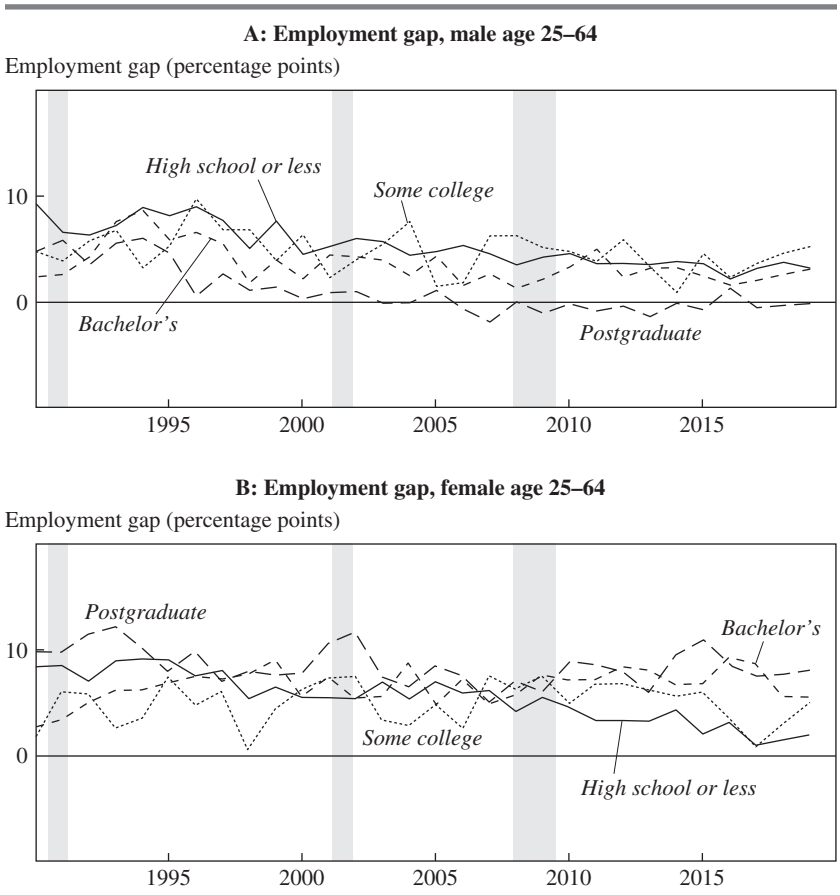


Source: Authors' calculations using CPS data.

are significantly different. Black and API+ women work more hours than their white or Hispanic counterparts. That said, during the last expansion, hours worked among white women grew somewhat, narrowing the gap. Hours for Hispanic women fell during the Great Recession and only started to climb again in 2017.

Previous authors have found similar gaps in a range of other labor market outcomes, including labor force participation, unemployment, and underemployment (Cajner and others 2017) and separation rates and job finding rates (Daly, Hobijn, and Pedtke 2017). Moreover, these authors show that the outcomes cannot be fully explained by differences in age, education,

Figure 4. White/API+ Trends in Employment Gaps by Education



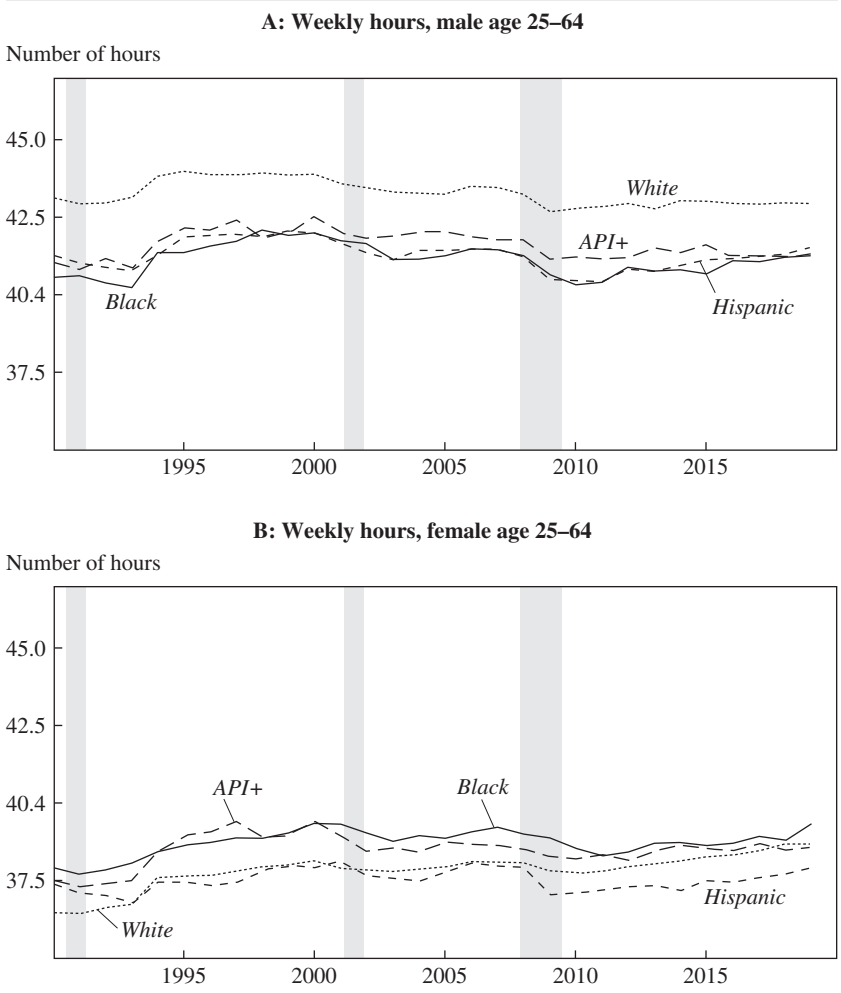
Source: Authors' calculations using CPS data.

experience, or occupation and industry but rather reflect unmeasured factors.[14] The persistent and large differentials in the employment of racial and ethnic minorities, even after controlling for educational attainment, point to considerable underutilized human resources that, if more equitably allocated, could boost aggregate output (Stewart and others 2021).

14. One well-studied unmeasured factor is discriminatory practices. Using specialized data and experiments, several authors have documented discriminatory practices or biases as an ongoing barrier in the labor market. See Neumark (2018), Altonji and Blank (1999), Bertrand and Mullainathan (2004), and Pager, Western, and Bonikowski (2009).

15985-02a_Buckman-6thPgs.indd 80 5/12/22 12:10 PM

Figure 5. Trends in Hours



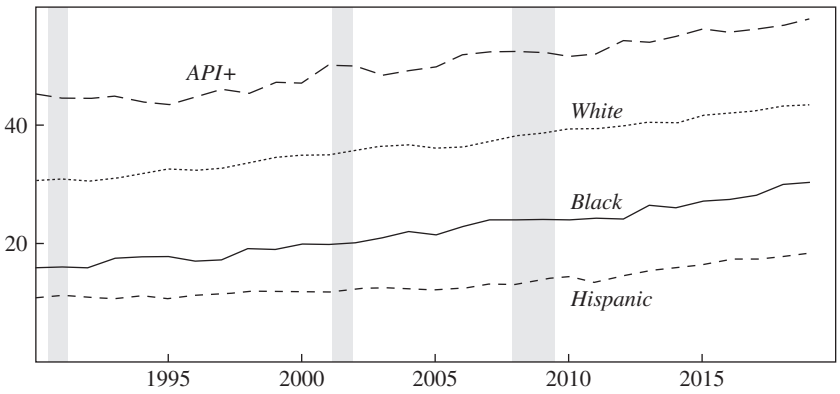
Source: Authors' calculations using CPS data.

Although educational attainment is not a remedy for all the gaps by race and ethnicity we observe, ongoing differences in education do play a role in determining differences in labor market outcomes. As such, we highlight them here. Figure 6 shows large and persistent differences in educational attainment across groups. The figure plots the share of men and women with a bachelor's degree or more by race and ethnicity from 1990 to 2019. Several things are worth noting. First, API+ have the largest share of the

Figure 6. Trends in Educational Attainment

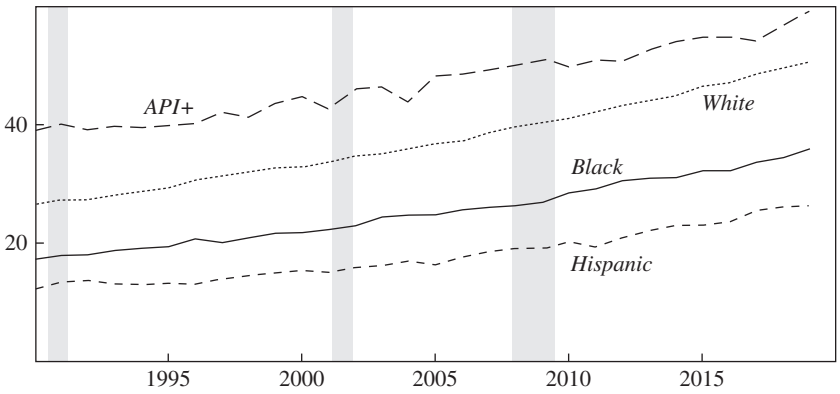
A: Share of population with a bachelor's degree or higher, male age 25–64

Percent share



B: Share of population with a bachelor's degree or higher, female age 25–64

Percent share



Source: Authors' calculations using CPS data.

population with a bachelor's or higher, driven largely by the very high college completion rates of Asian Americans, which is the largest subgroup of API+.[15] This holds for both men and women. White men and women have the next highest shares, followed by Black and then Hispanic men and women. These trends are consistent with findings by Espinosa and others (2019) on racial gaps in college completion.

15. Among Asian Americans, there is significant variation in the educational experiences of ethnic subgroups (Espinosa and others 2019).

15985-02a_Buckman-6thPgs.indd 82 5/12/22 12:10 PM

The differences across groups are large. For instance, in 2019, 58 percent of API+ men had a bachelor's degree or higher compared to 43.4 percent of white men, 30.4 percent of Black men, and 18.4 percent of Hispanic men. For women, the percentages were 59.1, 50.5, 35.9, and 26.3, respectively. These differences are larger than they were at the start of the sample for Black and Hispanic women as compared to white women. In fact, the gaps between white and Black women increased in the second half of the sample, as the rate of white women completing a bachelor's degree or higher increased, while the rate for Black women stayed relatively constant. Notably, we find these patterns are insensitive to the age range of the sample, suggesting that these gaps are not simply a reflection of past differences in educational access that have not yet aged out of the sample we study. Indeed, figure 7 shows that gaps in educational attainment are also large among those age 25–34, suggesting that differences in educational attainment continue to be an issue for racial and ethnic minorities.

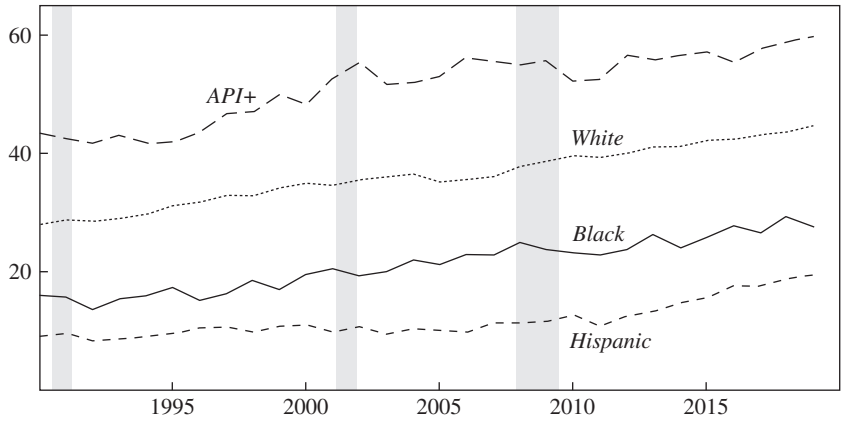
Even when gaps in educational attainment are closed, research has found that racial and ethnic minorities are not always in occupations consistent with their degrees (Williams and Wilson 2019; Abel and Deitz 2016). Figure 8 shows trends in educational utilization by group. Recall that being utilized is defined as being in an occupation that requires the level of education acquired. The figure plots the share of people with a bachelor's degree or more who are in an occupation requiring that level of education, by gender, race, and ethnicity. For both men and women, utilization rates of white and API+ workers are higher than those of Black and Hispanic workers. The gap is especially large for Black and Hispanic men. Black and Hispanic women have higher utilization rates than their male counterparts. Notably, the gaps in utilization for both genders have grown over time as white and API+ workers have become better utilized while utilization rates for Black and Hispanic workers have remained steady. In 2019, the last year of our sample, the utilization gap between white workers and both Black and Hispanic workers stood over 8 percentage points for both men and women.

The final trends we highlight are for earnings. Figure 9 shows real average hourly earnings for employed workers (in 2019 dollars). Starting with men, there is a sizable gap between the earnings of white and API+ workers and the earnings of Black and Hispanic workers. Over the sample, Black men have earned about 73 percent of what white men earn; Hispanic men earned about 68 percent of what white men earned. Although the gaps in earnings for females are smaller, they have notably widened over time. In 1990, the average Black female earned about 87 percent of what the

Figure 7. Trends in Educational Attainment

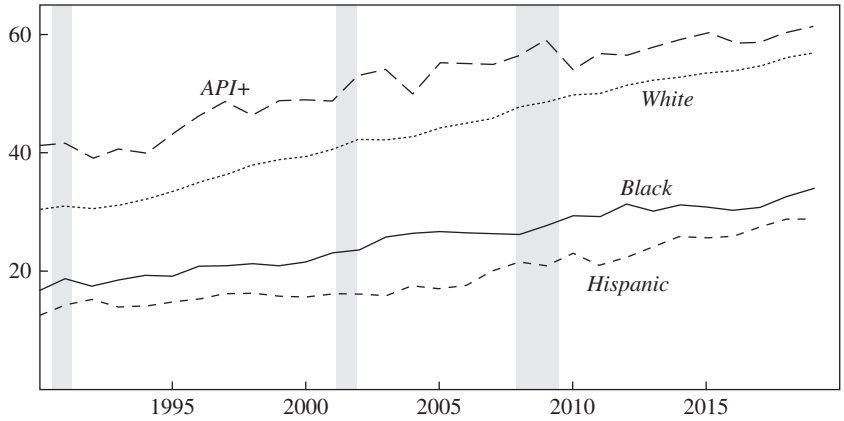
A: Share of population with a bachelor's degree or higher, male age 25–34

Percent share



B: Share of population with a bachelor's degree or higher, female age 25–34

Percent share



Source: Authors' calculations using CPS data.

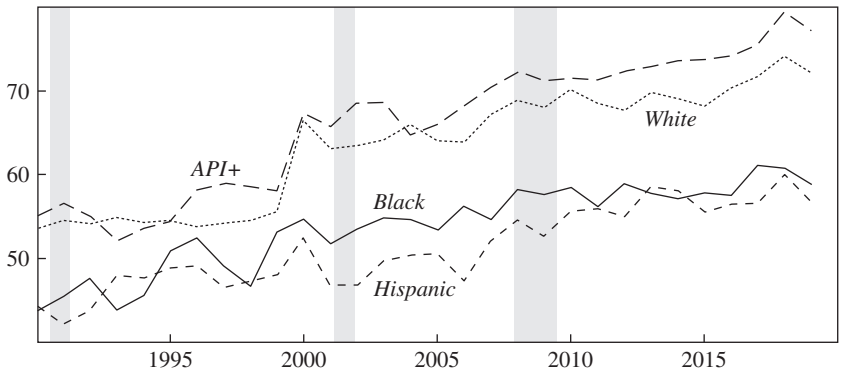
average white female earned. As of 2019, the average Black female earned only 82 percent of what the average white female earned. Similarly, the gap in earnings between white and Hispanic women has increased. In 1990 the average Hispanic female earned 81 percent of what the average white female earned; in 2019 this number had fallen to just 76 percent. These trends are consistent with findings from Wilson and Rodgers (2016).

15985-02a_Buckman-6thPgs.indd 84 5/12/22 12:10 PM

Figure 8. Trends in Educational Utilization (Age 25–64)

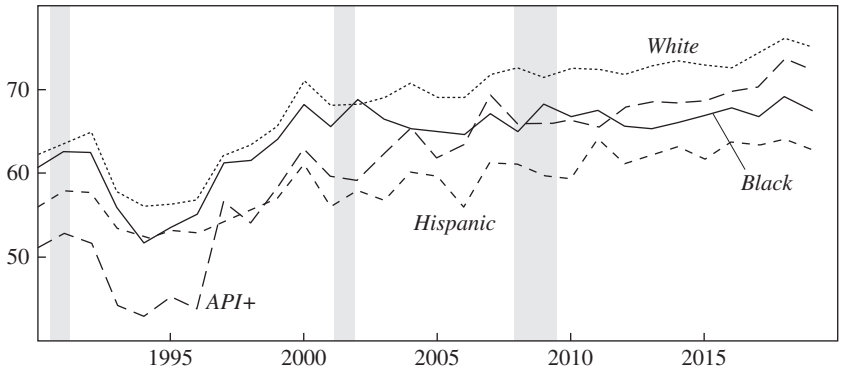
A: Share of those with bachelor's degree and postgraduates in high-skill jobs, male

Percent share



B: Share of those with bachelor's degree and postgraduates in high-skill jobs, female

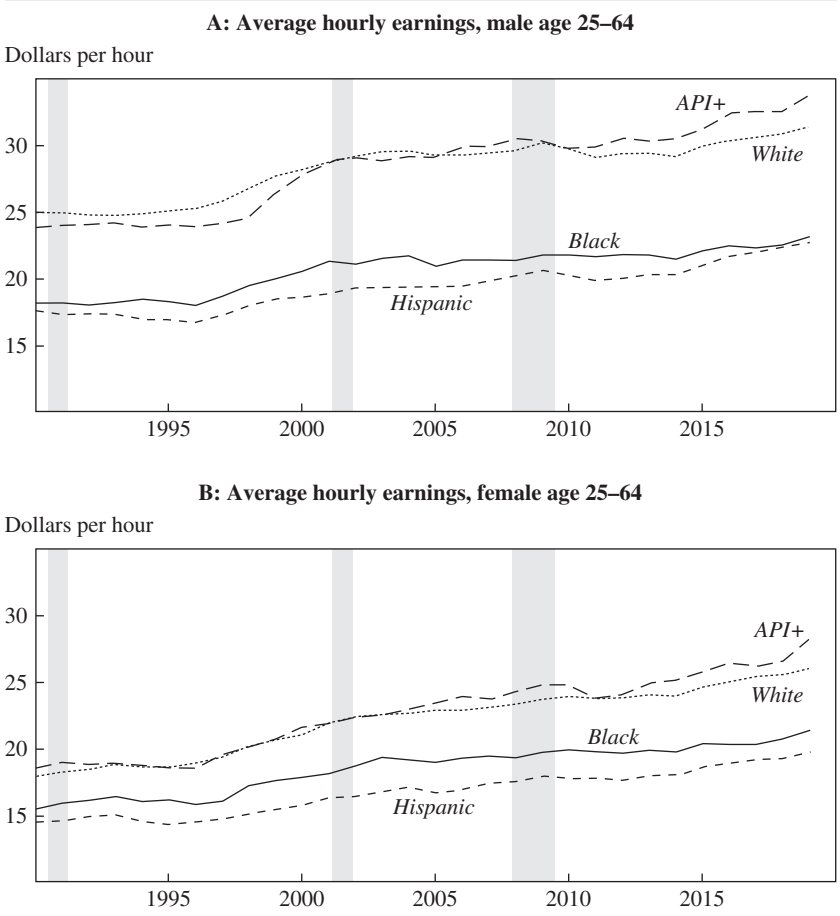
Percent share



Source: Authors' calculations using CPS data.

Of course, these earnings disparities reflect in part the differences in educational attainment and educational utilization highlighted previously. But as with employment gaps, differences in education cannot account for all of the earnings gaps. One way to see this is to look at earnings gaps by educational attainment, plotted in figures 10 and 11. The figures show the white/Black, white/Hispanic, and white/API+ earnings gaps by education, computed as white average hourly earnings less the average hourly earnings for other racial or ethnic groups. The results are striking. Among men, the white/Black earnings gap is consistently larger for those

Figure 9. Trends in Real Average Hourly Earnings in 2019 Dollars



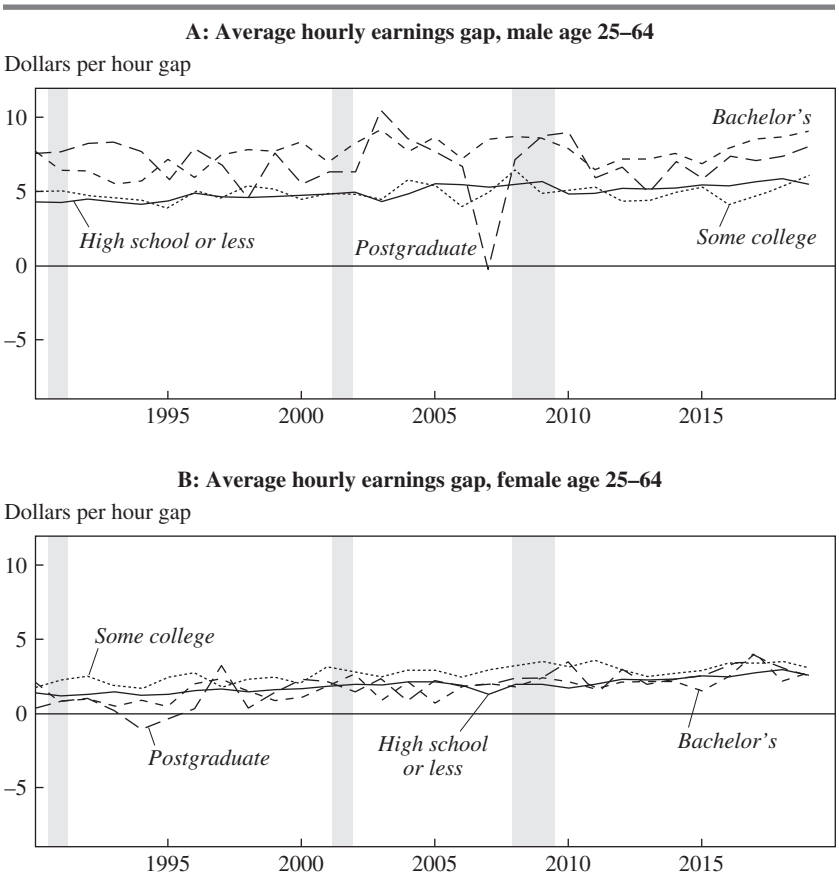
Source: Authors' calculations using CPS data.

with a bachelor's degree or higher, but sizable gaps exist for all education groups. Black women have a smaller earnings gap after controlling for education, but it has been steadily growing since 1990 for all education levels.

Although the earnings gaps for Hispanic men are somewhat smaller, gaps do exist for all levels of educational attainment. Similar to the pattern for Black men, increased education does not mean a smaller earnings gap. Hispanic men with a bachelor's or postgraduate degree have roughly the

15985-02a_Buckman-6thPgs.indd 86 5/12/22 12:10 PM

Figure 10. White/Black Trends in Real Average Hourly Earnings Gaps by Education in 2019 Dollars



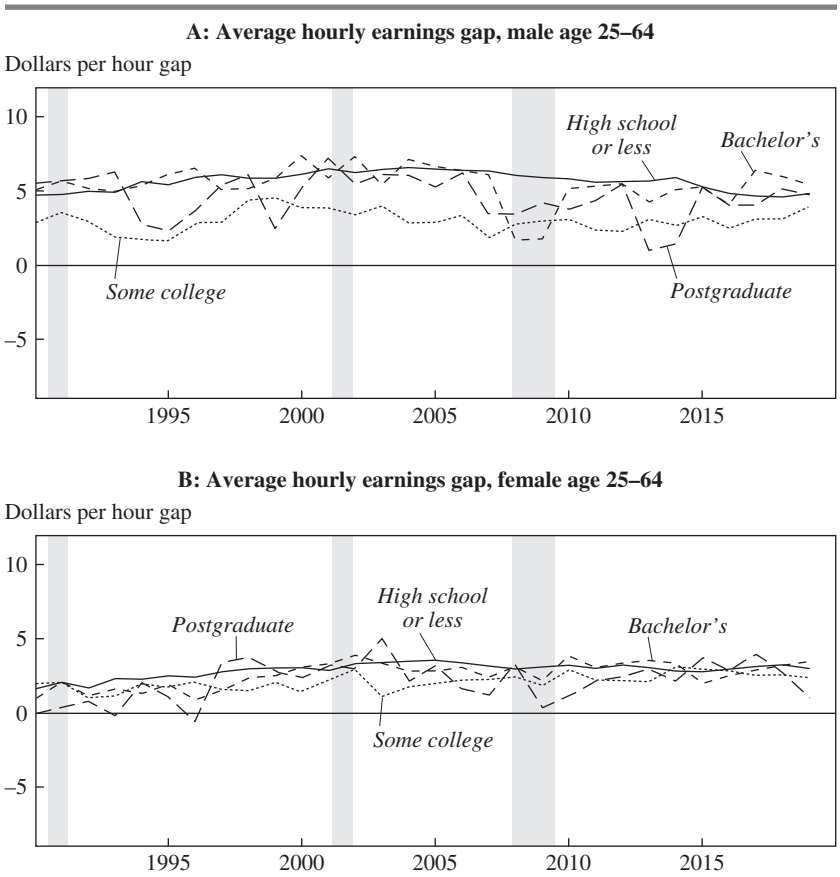
Source: Authors' calculations using CPS data.

same earnings gaps in 2019 as Hispanic men with a high school education or less. Hispanic women have even smaller earnings gaps, but like their male counterparts, there is little difference in the gaps by education.

Finally, figure 12 shows the patterns in earnings by education for API+ men and women. Among API+ men, those with less than a bachelor's degree have positive, although relatively small, earnings gaps compared to equally educated white men. In contrast, API+ men with a bachelor's degree or more earn more than their white counterparts on average, producing a negative earnings gap. This was not always the case, but it started to shift

15985-02a_Buckman-6thPgs.indd 87 5/12/22 12:10 PM

Figure 11. White/Hispanic Trends in Real Average Hourly Earnings Gaps by Education in 2019 Dollars



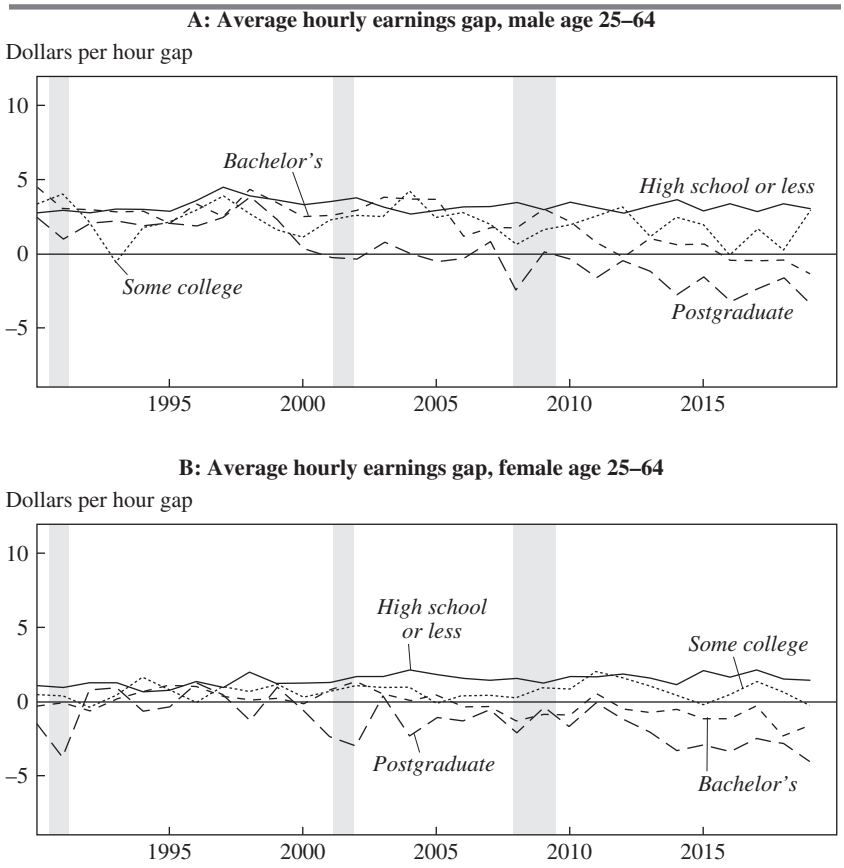
Source: Authors' calculations using CPS data.

in the 2000s for postgraduates and in the 2010s for those with a bachelor's degree. A similar pattern can be seen for API+ women. There are relatively small gaps for those with high school or less and negative gaps for those with a bachelor's degree or more.

Taken together, these trends document large and persistent gaps in labor market outcomes by race and ethnicity. These gaps reflect a variety of factors including ones that are easy and hard to measure. Whatever the cause, the disparities highlight inherently large gains from more equal outcomes. We turn to this now.

15985-02a_Buckman-6thPgs.indd 88 5/12/22 12:10 PM

Figure 12. White/API+ Trends in Real Average Hourly Earnings Gaps by Education



Source: Authors' calculations using CPS data.

II. Closing Labor Market Gaps—Previous Literature

Although most of the research on labor market disparities by race and ethnicity focuses on why gaps exist across groups, a recent set of studies has looked at the toll these disparities take on the economy or, said differently, how much better the economy would be doing if the gaps were erased. The findings point to large gains in GDP. For example, Hsieh and others (2019) examine the effect on aggregate productivity of the convergence in the occupational distribution between 1960 and 2010. They use a structural model to examine how much of the gain in productivity during this period

is associated with reductions in barriers to entry for women and Black workers. They estimate that the improved allocation of talent contributed between 20 percent and 40 percent of the total growth in aggregate market output per person during this period. They use a general equilibrium model to decompose the contribution of various forces, namely, discrimination, barriers to human capital formation, and differences in preferences or social norms. They find that lowering human capital barriers explains 36 percent of growth in GDP per person over the period, while declining labor market discrimination explains 8 percent of growth, and changing preferences explain little of the growth during this period.

Taking a nonstructural approach, Peterson and Mann (2020) conduct a simple empirical exercise to estimate the cost of Black inequality in the United States. They find that closing gaps between Black and white adults in wages, higher education, homeownership, and entrepreneurship would have generated significant additional income for saving, investing, and consumption, which would have led to a GDP boost of $16 trillion over the past twenty years and a projected $5 trillion gain over the next five years. Noel and others (2019) consider a similar question but focus on wealth. Through the lens of an Oxford model, they examine what closing the Black/white wealth gap by 2028 would do to aggregate GDP, using income, tangible investments, and stock market investments as components of wealth. They find that closing these gaps would increase aggregate output by 4–6 percent by 2028.

Other authors have considered the impact of closing more specific racial and ethnic gaps. For example, Turner (2018) focuses on closing the racial earnings gap associated with disparities in health, education, incarceration, and employment opportunities. The exercise sets earnings for minority groups, further divided by age and gender, to the average earnings of their non-Hispanic white counterparts. She finds that closing these gaps today would increase GDP by 22 percent by 2050, for a corresponding gain of $8 trillion. In a similar exercise, Treuhaft, Scoggins, and Tran (2014) estimate gains in average annual income and GDP under a hypothetical scenario in which there is no inequality of earnings or employment by race or ethnicity. The authors estimate the actual average annual income and hours of work for each racial and ethnic group, as well as projected values under the assumption that all racial and ethnic groups had the same average annual income and hours of work, by income percentile and age group, as the non-Hispanic white population. The projected values are then applied to the individual level for all racial and ethnic groups other than non-Hispanic white individuals. They find that closing these racial gaps in 2012 would have increased GDP by 14 percent or $2.1 trillion that year.

15985-02a_Buckman-6thPgs.indd 90 5/12/22 12:10 PM

Each of these empirical studies imagines a counterfactual world in which gaps by race and ethnicity do not exist and then computes the effect on aggregate GDP and some set of its components. Our paper adds to this literature. We make three contributions. First, we consider closing the gap for a wider span of racial and ethnic groups, including Black, Hispanic, and API+. Previous studies have tended to focus on closing gaps for a narrower set of demographic groups, such as Black people and women. Second, we perform our analysis annually for thirty years, from 1990 to 2019, and examine how the potential gains from greater labor market equity have changed over time. Finally, we investigate which outcome gaps among employment, hours, educational attainment, educational utilization, and residual earnings account for the GDP gains we find and whether these key drivers have remained constant or changed over time. These additions expand the scope of past work and provide additional insights into the areas policymakers might focus on to create positive change.

III. The GDP Gains from Eliminating Racial and Ethnic Gaps

The starting point for our analysis is the basic GDP math that expresses aggregate output, Y, as a function of physical capital, K, and labor input, L, in which we hold the relationship between capital and labor fixed:

$$Y = F(K, L). \tag{1}$$

The size of aggregate output depends on the amount of capital and labor used. This means that the gaps in labor market outcomes, like the ones shown in section I, translate directly into lower aggregate output.

The exercises that follow take this basic GDP math to the CPS data to examine how making the labor market outcomes for Black, Hispanic, and API+ individuals age 25–64 at least as favorable as their non-Hispanic white counterparts would have changed aggregate output in the United States over the past three decades.

III.A. Simple Counterfactual

We begin with a simple exercise that highlights the intuition of our experiments and ties us back to the results of previous research. Using data from the CPS for 2019, we select individuals age 25–64 who are not self-employed. We then compute the shares of this group who report being

Table 1. GDP Gain from Eliminating Minority Earnings Gaps in 2019

	Overall population age 25–64 (in millions)	*Group share of sample*	*Group average annual earnings*[a]	*Total group-specific earnings (in trillions)*
Observed labor earnings contribution to GDP				
White	155	0.60	46,397	4.31
Black	155	0.13	32,554	0.65
Hispanic	155	0.18	32,761	0.92
API+	155	0.09	47,430	0.68
Total	—	—	—	6.56
Counterfactual labor earnings contributions to GDP				
White	155	0.60	46,397	4.31
Black	155	0.13	46,397	0.93
Hispanic	155	0.18	46,397	1.30
API+	155	0.09	47,430	0.68
Total	—	—	—	7.21

a. This average is calculated inclusive of zeros, that is, if someone is unemployed, they are still included in the calculation and their earnings are set as zero.

Source: Authors' calculations using CPS data.

Note: Sample excludes self-employed individuals.

Black, Hispanic, API+, and white. Next, we compute average annual labor earnings by group, including those with and without positive earnings to account for differences in employment rates across groups. The data are shown in the top panel of table 1.

Multiplying the population total by that group's population share and by average annual earnings for each group yields the group-specific labor earnings contribution to total GDP. For example, for the white population, this calculation is 155 million $\times$ 0.60 $\times$ 46,397 = 4.31 trillion.

Using these numbers, we consider how much larger the labor earnings contribution to GDP would be if Black, Hispanic, and API+ workers had the same average earnings as white workers. Note that the earnings gaps capture the gaps in the labor market outcomes reviewed in section I as well as other unmeasured factors. This is a simple way to proxy a world in which all the labor market gaps by race and ethnicity that contribute to average earnings gaps are removed. As noted previously, we select the white population as our base group since they have long been the majority in the United States and have historically faced fewer systemic barriers in the labor market.

The results of the counterfactual exercise are reported in the bottom panel of table 1. Note that we do not change the earnings of the API+ group

since their average earnings are higher than white individuals.[16] This simple exercise makes an important point: eliminating gaps in average earnings by race and ethnicity, by bringing the average earnings of minorities to at least the level of the white average, would add notably to GDP, about $0.65 trillion in additional labor input in 2019, a 10 percent increase in labor income for the US economy.

Of course, this simple exercise, which relies on closing earnings gaps, falls short of capturing all the gains from closing other gaps in labor compensation in the economy. Assuming the earnings gains apply to all labor income, we can scale up our number using data from the National Income and Product Accounts (NIPA) data from the Bureau of Economic Analysis.[17] NIPA data tell us that aggregate employee compensation in the United States in 2019 was $11.45 trillion. Multiplying our percentage gain in labor income (10 percent) by aggregate employee compensation of $11.45 trillion generates an aggregate gain to GDP of about $1.15 trillion from eliminating gaps by race and ethnicity in 2019 alone. This number is consistent with the numbers generated by prior research, including Peterson and Mann (2020) and Treuhaft, Scoggins, and Tran (2014).

III.B. Changes in Gains over Time

Using this simple counterfactual exercise, we next ask whether the gains from more equal outcomes have risen or fallen over time. Two observations suggest that they might have risen. First, the trends we documented in section I show that while some labor market disparities by race and ethnicity have improved, others, especially for Black men and women, have remained stable or even worsened over time. Second, the population share of racial and ethnic minorities has been rising over the past three decades. This can be seen in table 2, which shows shares of the adult population age 25–64 who are not self-employed, by race and ethnicity. In 1990, 76 percent of this population was white, compared to 60 percent in 2019. Over the same period, the Hispanic share of the US population rose rapidly, increasing from 9 percent in 1990 to 18 percent in 2019. The API+ population share also rose and accounted for 9 percent of the US population in 2019. The share of the Black population remained relatively constant. Overall, this

16. This is a practice we maintain throughout the paper. Given the significant heterogeneity within the API+ group (Kochhar and Cilluffo 2018), our results for this group should be interpreted with caution.

17. Bureau of Economic Analysis, "Personal Income," table 6.2D; https://www.bea.gov/data/income-saving/personal-income.

Table 2. Share of Racial and Ethnic Minorities in US Population (Age 25–64)

	1990	*2000*	*2010*	*2019*
White	0.76	0.70	0.66	0.60
Black	0.12	0.12	0.12	0.13
Hispanic	0.09	0.12	0.15	0.18
API+	0.04	0.06	0.07	0.09

Source: Authors' calculations using CPS data.
Note: Sample excludes self-employed individuals.

Table 3. Changes over Time in GDP Gain from Eliminating Minority Earnings Gaps

	Level gains (2019 $ in trillions)				*Percent gains*			
	1990	*2000*	*2010*	*2019*	*1990*	*2000*	*2010*	*2019*
White	—	—	—	—	—	—	—	—
Black	0.15	0.19	0.25	0.28	4.04	3.80	4.56	4.21
Hispanic	0.11	0.22	0.31	0.38	3.06	4.46	5.67	5.81
API+	0.02	0.02	0.02	—	0.45	0.31	0.35	—
Total	0.28	0.43	0.58	0.66	7.55	8.57	10.58	10.02

Source: Authors' calculations using CPS data.
Note: Sample excludes self-employed individuals.

suggests that the labor market gaps described in section I affect an increasing share of the US population.

Table 3 shows how these facts have affected the gains from equalizing outcomes as computed in table 1 over time. The table reports gains to group-specific earnings and total economy-wide labor earnings from bringing average earnings for minorities to at least the level of the non-Hispanic white population for 1990, 2000, 2010, and 2019. The gains are the difference between observed group-specific labor earnings each year and the counterfactual group-specific labor earnings.[18]

Beginning with the last line of the table, total GDP gains from eliminating gaps in earnings have grown considerably over time, rising from $0.28 trillion in 2019 dollars in 1990 to $0.66 trillion in 2019. Note that the value of $0.66 trillion in 2019 is the difference of the observed and counterfactual group-specific labor earnings contributions in table 1, $7.21 trillion–$6.56 trillion. Since these numbers might be rising simply because the economy has grown, we also show percentage gains to GDP in the right panel of table 3. In percentage terms, counterfactual gains from labor income have risen from about 7.6 percent in 1990 to 10 percent in 2019.

18. As in the previous exercise, if average earnings for a group are higher than for non-Hispanic white workers, we do not change them.

15985-02a_Buckman-6thPgs.indd 94 5/12/22 12:10 PM

The patterns by race and ethnicity are also informative and capture the changes in trends in labor market gaps and population shares we discussed. For example, for Black individuals, the percentage gains from equalizing earnings to those of white individuals have remained relatively steady, boosting the labor earnings contribution to GDP by between 3.8 and 4.6 percent over the sample period. The percentage gains from equalizing earnings of Hispanic workers and non-Hispanic white workers has grown over time, owing largely to the rising share of the Hispanic population. For the API+ group, the contributions from closing gaps with the non-Hispanic white population have fallen as that group has experienced improved outcomes over time. However, as mentioned previously, this is a very heterogeneous group, so the average experience does not reflect the outcomes of many.

These simple counterfactual exercises show the potential gains to aggregate labor earnings and aggregate labor output if average earnings for racial and ethnic minorities were at least as high as white individuals. The results point to significant boosts to GDP that have been rising over time in both absolute and percentage gains. We now consider the factors contributing to these potential gains.

III.C. Drivers of the Gains

The exercises so far show the inherently large gains to GDP from eliminating gaps in average annual earnings by making racial and ethnic minorities at least as well off as the average white individual. We now turn to understanding the key drivers of the gains we've computed. As shown in section I there are many disparities in labor market outcomes that add up to differences in average earnings across groups. To identify the specific contributions of these disparities, we perform a more detailed counterfactual analysis that sequentially eliminates gaps in employment, hours, educational attainment, and educational utilization. Previous research has shown that these factors are important determinants of earnings differentials by race and ethnicity.[19]

19. See, for example, Daly, Hobijn, and Pedtke (2017), Altonji and Blank (1999), and O'Neill (1990). We recognize that other measurable factors also matter, such as industry and occupation (Matthews and Wilson 2018; Daly, Hobijn, and Pedtke 2017; Del Rio and Alonso-Villar 2015) and geographic location (Cajner and others 2017; Parks 2012), as well as a host of other factors that are more difficult to measure, including differences in educational quality (Card and Krueger 1992), differences in career ladder opportunities (Daly, Hobijn, and Pedtke 2020), and discrimination, current or historical (Darity and Mason 1998; Daly, Hobijn, and Pedtke 2017; Cajner and others 2017). Although we do not separately quantify them in this analysis, their effects are accounted for in the remaining differences in average earnings once differences in employment, hours, education, and educational utilization have been eliminated.

Our more detailed counterfactual exercises imagine a world in which these gaps are eliminated one by one and compute the gains to GDP associated with each. Recognizing the heterogeneity of outcomes by age and gender, we also disaggregate the data by age and gender.

DATA AND METHODS Before we can perform our more detailed counterfactual analysis, we need to make several adjustments to the data. Beginning with the same CPS sample used in the simple counterfactual, which covers civilian noninstitutionalized adults age 25–64 who are not self-employed, we collapse the data into groups defined by thirty-two mutually exclusive age, gender, and race/ethnicity cells: two genders (male and female), four race/ethnicity groups (white, Black, Hispanic, and API+), and four ten-year age ranges (25–34, 35–44, 45–54, and 55–64). For each of these thirty-two groups, we then compute group-specific values for our analysis variables: employment, hours, educational attainment, educational utilization, and earnings.[20]

Using these data and equation (2), we can recover the observed labor earnings contribution to GDP shown in table 1 as well as the group-specific contributions by race and ethnicity:

$$Y_L = \sum_r Y_{Lr} = \sum_r \left(P * \alpha_r\right) * w_r. \tag{2}$$

Y_L is the group-specific labor contribution to GDP, P is the population age 25–64 who are not self-employed, α_r is the population share of a given race/ethnicity group, and w_r is the group-specific average earnings. The labor contribution to GDP is the sum of the group-specific contributions. Substituting the white earnings value for racial and ethnic groups with negative earnings gaps, $w_{white} > w_{own}$, as in equation (3), produces the simple counterfactual (SC) shown in table 1:

$$Y_L^{SC} = \sum_r Y_{Lr} = \sum_r \left(P * \alpha_r\right) * w_r^{SC}. \tag{3}$$

Following this same logic, it is straightforward to perform a more detailed counterfactual exercise that adjusts the drivers of the observed earnings gaps of the employed, including employment, hours worked, education, and education utilization. And we can do this over more narrowly defined

20. These variables are defined as in section I. We adjust all values to reflect annual totals.

groups that separate the data by age, gender, and race/ethnicity. Equation (4) shows how this is done:

$$Y_L = \sum_{a,g,r} Y_{L_{a,g,r}} = \sum_{a,g,r} \left(P * \alpha_{a,g,r} \right) * e_{a,g,r} * h_{a,g,r} * w_{a,g,r}, \tag{4}$$

where $Y_{L_{a,g,r}}$ is the group-specific labor contribution to GDP, P is the population age 25–64 who are not self-employed, $\alpha_{a,g,r}$ is the population share of a given age/gender/race/ethnicity group, $e_{a,g,r}$ is the group-specific employment rate, $h_{a,g,r}$ is the group-specific average hours, and $w_{a,g,r}$ is the group-specific average hourly wage. The counterfactual labor contribution to GDP, $Y^C_{L_{a,g,r}}$, is the sum of the age/gender/race/ethnicity group-specific contributions accounting for adjustments to the underlying variables. The specific steps of this calculation are as follows.

Step 1: Adjust employment rates Calculate the counterfactual employment rate as the white employment rate for each age/gender group:

$$e^C_{a,g,r} = e_{a,g,white}. \tag{5}$$

Substitute the counterfactual employment rate for the observed employment rate in all cases where

$$e^C_{a,g,r} > e_{a,g,own}. \tag{6}$$

Recompute labor input with the adjusted employment rate, holding hours, education, education utilization, and average earnings at the own group-specific values:

$$Y^C_{L_{a,g,r}} = \left(P * \alpha_{a,g,r} \right) * e^C_{a,g,r} * h_{a,g,r} * w_{a,g,r}. \tag{7}$$

The difference between the employment-adjusted counterfactual contribution and the observed contribution of labor earnings to GDP is the GDP effect of closing employment gaps by race and ethnicity. The next step is to add the hours adjustment.

Step 2: Average weekly hours Conditional on employment, calculate the counterfactual average weekly hours as the non-Hispanic white hours for each age/gender group:

$$h^C_{a,g,r} = h_{a,g,white}. \tag{8}$$

Substitute the counterfactual hours for the observed hours in all cases where

$$h^{C}_{a,g,r} > h_{a,g,own}. \tag{9}$$

Recompute labor input with adjusted employment rate and hours, holding education, education utilization, and average earnings at the own group-specific values:

$$Y^{C}_{L_{a,g,r}} = \left(P * \alpha_{a,g,r}\right) * e^{C}_{a,g,r} * h^{C}_{a,g,r} * w_{a,g,r}. \tag{10}$$

The difference between the employment-adjusted counterfactual contribution and the employment- and hours-adjusted counterfactual contribution is the additional GDP effect of closing hours gaps by race and ethnicity. We next turn to education and education utilization.

Step 3: Educational attainment Adjustments to education and education utilization affect the return on labor input, or average earnings, $w_{a,g,r}$. So we begin by adjusting the distribution of education by race and ethnicity for each of our age/gender cells using the following steps.

Conditional on employment, shift the whole vector of education (*ed*) for a race/ethnicity group if the sum of the shares of those with some college, a bachelor's degree, or a postgraduate degree is less than the non-Hispanic white share for that age/gender group:

$$ed^{C}_{a,g,r} = ed_{a,g,white}. \tag{11}$$

Substitute the counterfactual educational distribution for the observed educational distribution in all cases where

$$ed^{C}_{a,g,r} > ed_{a,g,own}. \tag{12}$$

Recompute labor input with adjusted employment rate, hours, and education, holding education utilization, and average earnings at the own group wage for an age/gender/race/ethnicity group conditional on being employed.

Equalizing the distribution in education by race and ethnicity so that it is at least at the level of the white population, boosts the average wage for the affected groups, $w^{C'}_{a,g,r}$:

$$Y^{C}_{L_{a,g,r}} = \left(P * \alpha_{a,g,r}\right) * e^{C}_{a,g,r} * h^{C}_{a,g,r} * w^{C'}_{a,g,r}. \tag{13}$$

The difference between the employment- and hours-adjusted counterfactual contribution and the employment-, hours-, and education-adjusted counterfactual is the additional GDP effect of closing gaps in the education distribution by race and ethnicity. The next step is to add in utilization.

Step 4: Education utilization Conditional on employment, calculate the counterfactual utilization distribution (u) as the white utilization:

$$u^{C}_{a,g,r} = u_{a,g,white}. \tag{14}$$

Substitute the counterfactual utilization distribution for the observed utilization distribution in all cases where

$$u^{C}_{a,g,r} > u_{a,g,own}. \tag{15}$$

Recompute labor input with the adjusted employment rate, hours, education, and education utilization, holding average earnings at the group-specific wage conditional on employment. Like the adjustment to the distribution of education, changing education utilization for racial and ethnic minorities so that they are at least as well utilized as white individuals boosts average earnings for the group, expressed as $w^{C''}_{a,g,r}$:

$$Y^{C}_{L_{a,g,r}} = \left(P * \alpha_{a,g,r}\right) * e^{C}_{a,g,r} * h^{C}_{a,g,r} * w^{C''}_{a,g,r}. \tag{16}$$

The difference between the employment-, hours-, and education-adjusted counterfactual contribution and the employment-, hours-, education-, and education utilization–adjusted counterfactual contribution is the additional GDP effect of closing gaps in the utilization distribution by race and ethnicity.

Step 5: Residual earnings The final step in the process is to calculate the contribution of residual average earnings gaps that are not explained by differences in employment, hours, education, and education utilization. We compute this residual as

$$w^{res} = Y^{SC}_{L} - Y^{C}_{L}. \tag{17}$$

This difference is the contribution to total gains made by factors not accounted for in our analysis. These could be other observable factors, such as industry and occupation, as well as less observable ones, such as discrimination or other barriers.

Table 4. GDP Gains and Percentage Contribution from Eliminating Minority Labor Market Gaps

	Employment rate	*Hours*	*Education*	*Utilization*	w^{res}	*Total CPS*	*Total NIPA*
Labor earnings contribution—level gains in trillions of 2019 dollars							
1990	0.07	0.02	0.04	0.00	0.14	0.27	0.46
2019	0.15	0.06	0.19	0.04	0.29	0.73	1.28
Labor earnings contribution—percentage gains							
1990	2.1	0.7	1.2	0.1	3.8	7.9	—
2019	2.2	0.9	2.9	0.6	4.5	11.1	—
Share contribution to GDP							
1990	26.5	8.6	15.2	1.7	48.0		
2019	20.0	8.3	26.3	5.6	39.8		

Source: Authors' calculations using CPS data.
Note: Sample excludes self-employed individuals.

III.D. Main Results

The results of our more detailed counterfactual exercise are displayed in table 4. For simplicity the table shows results for 1990 and 2019.[21] The first thing to note is that the estimated gains from our more detailed counterfactual are slightly larger in each year than in the simple counterfactual. For example, in 2019 the estimated gains from the detailed counterfactual are \$0.73 trillion, as compared to \$0.65 trillion in the simple counterfactual. This reflects the additional disaggregation by age and gender incorporated into this exercise, primarily the significant heterogeneity by gender reported in section I, although age also matters. Importantly, as we will show, the additional disaggregation does not affect the patterns over time, just the levels and percentage gains relative to observed GDP.

The second thing to note is that, as in the simple counterfactual results reported in table 3, the gains from closing gaps by race and ethnicity have risen over time, in both absolute and percentage terms. This can be seen in the second-to-last column of the first two panels of table 4. Looking at the level gains, closing labor market gaps related to earnings of non-self-employed individuals age 25–64 in 1990 would have resulted in \$0.27 trillion of additional GDP. Scaling this to apply to all labor compensation for all workers, using the NIPA data, the total gain to GDP would have been \$0.46 trillion, as shown in the final column of the first two panels. By 2019, these values had risen to \$0.73 trillion and \$1.28 trillion, respectively.

21. The results for the other years in our sample follow the trends reported in table 3.

15985-02a_Buckman-6thPgs.indd 100 5/12/22 12:10 PM

Turning to the drivers of the gains, as seen in the bottom panel of table 4, the results show that employment and educational attainment make the largest contributions among the analysis variables, although their relative importance has shifted over time.[22] For example, in 1990, differences in employment were more important than differences in education. The reverse was true in 2019. Turning to hours and utilization, they matter but their effects are small relative to employment and education. That said, over time, educational utilization, like educational attainment, has become an increasingly important contributor to the potential gains from greater labor market equity. The importance of hours has stayed relatively constant over the period.

Finally, consistent with prior research, residual earnings gaps, w^{res}, play a large role in generating the overall gains to GDP we compute. This is consistent with prior research that finds significant earnings gaps remain for racial and ethnic minorities even when controlling for a large number of explanatory factors (Daly, Hobijn, and Pedtke 2017; Cajner and others 2017). In our analysis, residual earnings gaps account for about 48 percent of the gains in 1990 and 40 percent in 2019.

DIFFERENCE IN DRIVERS ACROSS RACE/ETHNICITY AND GENDER The main results presented in the prior sections highlight the important role that employment, education, and residual earnings play in explaining the total gains to GDP that we calculate in our simple counterfactual. In this section, we further examine the relative importance that these factors play within each racial or ethnic minority group. We also examine whether these overall patterns differ when considering both race/ethnicity and gender. Table 5 shows the percentage contribution to the group-specific GDP gains of each of our key drivers: employment, hours, education, utilization, and w^{res}, for our three race/ethnicity groups in 1990 and 2019. The sum across the columns equals 100, with rounding error. The first panel shows the contributions for the total sample, which is a replication of the last panel in table 4, repeated for convenience of comparison across groups. The next three panels repeat the exercise for Black, Hispanic, and API+ groups, inclusive of both men and women.

Comparing the last three panels, it is immediately clear that the drivers of the overall gains differ considerably by race and ethnicity. For Black individuals, education plays less of a role and residual earnings gaps play

22. When comparing these findings to prior research it is important to keep in mind that we have made sequential adjustments, meaning that as we move across the table the values show the additional gains to GDP from adjusting each variable.

15985-02a_Buckman-6thPgs.indd 101 5/12/22 12:10 PM

Table 5. Share Contribution to Group-Specific GDP Gains by Race

	Employment rate	*Hours*	*Education*	*Utilization*	w^{res}
All					
1990	26.5	8.6	15.2	1.7	48.0
2019	20.0	8.3	26.3	5.6	39.8
Black					
1990	23.6	8.6	16.9	1.2	49.6
2019	21.8	5.5	18.2	4.8	49.6
Hispanic					
1990	22.3	6.0	29.0	1.5	41.2
2019	12.8	6.7	41.2	6.9	32.4
API+					
1990	41.9	13.3	–1.1	3.7	42.2
2019	46.4	24.4	–5.8	4.4	30.5

Source: Authors' calculations using CPS data.
Note: Sample excludes self-employed individuals.

a much larger role. This is true in both 1990 and 2019. Notably, for Black individuals there is no change in the role that residual earnings gaps play over time, consistent with the literature that finds little progress in reducing residual earnings gaps over the sample period (Daly, Hobijn, and Pedtke 2017; Cajner and others 2017). For Hispanic individuals, the role of gaps in educational attainment and educational utilization increased notably between 1990 and 2019, as the role of employment fell. In 2019, closing gaps in education, once gaps in employment and hours were eliminated, accounted for 41 percent of the overall gains to GDP from the Hispanic group. Finally, for API+, closing gaps in employment are the primary driver of gains, accounting for 42 percent of the total gains in 1990 and 46 percent in 2019. Hours also matter more for API+ than for other groups.

Our final table further disaggregates the results to consider race and gender. The trends in section I pointed to large differences by gender in many of the key variables. Table 6 repeats the analysis in table 5 by race and gender. Beginning with the results for Black men and women, a couple of differences stand out. First, employment gaps and hours gaps play a larger role for Black men than Black women, especially in 2019. Second, gaps in educational attainment, adjusting for employment and hours, are significantly more important for Black women than for Black men. For both groups, closing residual earnings gaps explains about half of the total computed gains to GDP.

15985-02a_Buckman-6thPgs.indd 102 5/12/22 12:10 PM

Table 6. Share Contribution to Group-Specific GDP Gains by Race and Gender

	Employment rate	*Hours*	*Education*	*Utilization*	w^{res}
Black males					
1990	22.8	11.2	14.4	1.3	50.3
2019	26.0	7.8	11.4	4.2	50.7
Black females					
1990	26.2	0.0	25.4	0.9	47.5
2019	11.5	0.0	35.2	6.6	46.8
Hispanic males					
1990	9.6	9.4	29.6	1.6	49.7
2019	2.1	8.3	41.5	7.6	40.5
Hispanic females					
1990	44.9	0.0	27.9	1.3	25.9
2019	26.4	4.8	40.7	6.0	22.0
API+ males					
1990	30.2	19.5	–0.7	1.8	49.4
2019	28.7	35.5	–3.3	3.3	35.8
API+ females					
1990	66.8	0.0	–1.9	7.8	27.2
2019	73.1	7.4	–9.1	6.1	22.5

Source: Authors' calculations using CPS data.
Note: Sample excludes self-employed individuals.

There are even larger differences between Hispanic men and women when it comes to employment and residual earnings gaps. Employment gaps play a much larger role for Hispanic women than Hispanic men. In contrast, residual earnings gaps play a larger role for Hispanic men compared to Hispanic women. For both Hispanic men and women, closing educational attainment gaps accounts for the largest share of gains to GDP in 2019.

The last two panels show results for API+ men and women. Among API+ men, differences in hours play an outsized role relative to other groups. Moreover, the importance of hours has risen over time. Residual earnings gaps also play an important role, although their contribution has fallen over time. For API+ women, gaps in employment rates are by far the largest contributor, explaining about 67 percent of the gains in 1990 and 73 percent in 2019. The importance of hours gaps has increased over time but remains small relative to API+ men. Residual earnings gaps matter, but similar to Hispanic women, are small relative to most other groups.

The results of this disaggregation highlight an important point. When studying economic outcomes by race and ethnicity it is critical to look

beneath simple averages. There are significant differences in experiences by race and ethnicity and by gender that make it vital to be as detailed as the data will permit. Building a complete picture of the gaps across these more narrowly defined groups is vital for shaping our understanding and for building the policies needed to resolve them.

III.E Discussion and Caveats

Our main results show how closing gaps in labor market variables by race and ethnicity, moving all groups to at least the levels of the white population, would boost the overall level of GDP in the nation. The gains are large and point to significant increases in overall economic output from greater labor market equity: $0.73 trillion in labor earnings in 2019 which, when scaled to match total labor compensation, amount to $1.28 trillion. Our results also show that the gains from equity have risen over time, as the fraction of racial and ethnic minorities in the population has increased. Adding up the counterfactual contributions over the thirty-year period of our sample, 1990 through 2019, totals $15.2 trillion in labor earnings, and $25.6 trillion in total labor compensation.

Given that the population share of racial and ethnic minorities is expected to rise further in the future, the gains from equity will continue to grow. Using population projections from the Census Bureau and our 2019 contribution values, we estimate total gains to labor earnings from closing racial and ethnic labor market gaps.[23] We project gains in labor earnings of $0.9 trillion in 2030, compared to $0.73 in 2019, which would scale to even higher values of total labor compensation.

Of course, our calculation does not take into account any of the general equilibrium effects of making these changes. One important general equilibrium effect is that the overall wages could adjust in response to higher employment rates and greater hours worked of racial and ethnic minorities. Our base case assumes that eliminating employment and hours gaps by race and ethnicity has no effect on wages—the economy just scales up to absorb the new labor supply. But theory tells us that some adjustments are likely. Although there is not much recent research on this topic for the overall labor market, there is literature on the impact of immigration on the wages of native workers. In general this literature finds small effects, with wage elasticities ranging from 0.1 to –0.1 (Peri 2014).[24] Applying the largest adjustment

23. US Census Bureau, "Data," https://www.census.gov/programs-surveys/popproj/data/datasets/2017/demo/popproj/2017-popproj.html. We use the main series from data set 1.

24. Peri (2014) reviews twenty-seven empirical studies and finds that two-thirds of them estimate wage elasticities between –0.1 and 0.1.

15985-02a_Buckman-6thPgs.indd 104 5/12/22 12:10 PM

(–0.1) to our results reduces our estimated gains to labor earnings in 2019 from $0.73 trillion to $0.71 trillion. Over the thirty years of our sample, the gains to labor earnings would fall from $15.2 trillion to $14.7 trillion, with total labor compensation falling from $25.6 trillion to $24.8 trillion.

So far, we have discussed why the results might be lower than our estimate. But they could also be higher. Research has shown that the direct gains to greater equity in the labor market are only the beginning. More equitable allocation of talent by education, employment, and jobs improves innovation, invention, and entrepreneurship, which set the foundation for growth today and growth in the future (Bell and others 2019; Aghion and others 2018). And a growing body of research finds that including more people in the economy from different backgrounds allows for more diverse teams, which contributes to better performance. For example, Kline, Rose, and Walters (2021) find that racially discriminatory hiring practices among firms are negatively correlated with firm profitability, while Herring (2009) finds that among for-profit business organizations, racial diversity in the workforce is associated with positive performance indicators like increased sales revenue, greater market share, and greater relative profits. This type of research suggests that investments and actions by private sector businesses to close labor market gaps can directly benefit their bottom line. All of these impacts are outside of our baseline calculations and could boost the gains beyond what we have measured.

IV. Conclusion and Future Research

The opportunity to participate in the economy and to succeed based on ability and effort is at the foundation of our nation and our economy. Unfortunately, structural barriers have persistently disrupted this narrative for many Americans, leaving the talents of millions of people underutilized or on the sidelines. The result is lower prosperity, not just for those affected, but for the economy as a whole.

Here we have put forth some initial estimates of the economic gains from equalizing outcomes for minorities to at least the level of the non-Hispanic white population. The gains are large and persistent and likely to increase further in the future. So what is holding us back? If the gains to equity are so large, why haven't we been able to close the gaps?

The answers often lie beyond the economics literature, and a large body of multidisciplinary research points to some of the hurdles. For one, many of the structural barriers we see have become deeply embedded in our society and economy, the result of historic discriminatory policies and practices,

such as Jim Crow laws and redlining, which have left enduring impacts on many racial and ethnic minorities (Rothstein 2017; Oliver and Shapiro 1995; Denton and Massey 1993).[25] The accumulation of these inequities over multiple decades and generations suggests that achieving equity will take time and significant investment. It won't be as easy as declaring them gone.

This is clear when looking at the long history of trying to reduce barriers in educational attainment. Eliminating the differences in human capital investment by race and ethnicity will likely take a large influx of resources over a number of years and the returns on investment likely won't accrue until well into the future. But when they are made, the evidence shows they pay off. A good example is the return on high-quality early childhood education programs. García and others (2020) find that such programs improve educational attainment and labor income later in life. In their particular study, they documented a 13 percent annual rate of return, net of the cost of financing the programs. Of course, closing gaps in labor market income also contributes to closing racial wealth gaps (Aliprantis and Carroll 2019). And this is important for future gains, producing the positive cycle of "wealth begets wealth" (Black and others 2020) that has been so important to intergenerational success.

Investments in reducing other gaps will also take up-front investment and likely produce only downstream benefits. This means that future research needs to focus on multiyear analysis that enables policymakers to evaluate the costs and benefits for cohorts and generations rather than just annually, where the investments don't always pencil out. It will also be important for future research to be interdisciplinary, combining the economics of closing gaps with the social and community benefits of improved equity.

All of this work will call for a new mindset—a mindset that sees large gaps by race and ethnicity as inefficient and a sign of misallocation, rather than an unfortunate, but efficient, outcome of a well-functioning market. Williams and Spriggs (1999), Mason and Williams (1997), and Darity and Williams (1985), among others, have noted that prevailing economic models, which assume efficient market outcomes, are challenged to deliver results that highlight the economic losses associated with the existence and persistence of labor market and other disparities.[26] Their point, as we read

25. Other examples that have been cited in the literature are governance structures imposed on Native Americans (Dippel 2014) and the internment of Japanese Americans (Chin 2005).

26. This point was also made by William Spriggs and Sendhil Mullainathan during Racism and the Economy: Focus on the Economics Profession, a virtual session hosted by all twelve district banks of the Federal Reserve System on April 13, 2021.

15985-02a_Buckman-6thPgs.indd 106 5/12/22 12:10 PM

it, is that if economic models start with the assumption that markets work perfectly, it naturally follows that persistent barriers, which limit productivity, would be removed by profit-seeking entrepreneurs. This then implies that remaining gaps are best explained by differences in productivity, even if hard to measure, or differences in group-specific preferences that drive people to sort by race and ethnicity into specific types or intensities of work. A long literature on persistent gaps follows this model.

But what if we started from a different vantage point and asserted that talent and innate preferences were distributed equally across racial and ethnic groups? This would naturally suggest that disparate outcomes were a misallocation and open the door for researchers to investigate how they might have arisen, including explanations related to historical and current differentials in investment, access, and labor market treatment. With considerable pressures weighing on US economic potential in coming decades, the time seems right to take a new perspective and imagine what's possible if equity is achieved.

ACKNOWLEDGMENTS We thank Katharine Abraham, John Fernald, Nicole Fortin, Erik Hurst, Huiyu Li, William Spriggs, and participants in the *BPEA* September 2021 conference for invaluable feedback on the project. This paper also benefited from comments and suggestions by seminar participants at the Federal Reserve Bank of San Francisco. Research assistance was provided by Olivia Lofton. The views expressed are those of the authors and do not necessarily represent the views of the Federal Reserve System or the Federal Reserve Bank of San Francisco. All remaining errors are our own.

References

Aaronson, Stephanie, Mary C. Daly, William Wascher, and David W. Wilcox. 2019. "Okun Revisited: Who Benefits Most from a Strong Economy?" *Brookings Papers on Economic Activity*, Spring, 333–75.

Abel, Jaison R., and Richard Deitz. 2016. "Underemployment in the Early Careers of College Graduates following the Great Recession." Working Paper 22654. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w22654.

Aghion, Phillipe, Ufuk Akcigit, Ari Hyytinen, and Otto Toivanen. 2018. "What Makes an Inventor?" Project Syndicate, January 6. https://www.project-syndicate.org/onpoint/what-makes-an-inventor-by-philippe-aghion-et-al-2018-01.

Aliprantis, Dionissi, and Daniel R. Carroll. 2019. "What Is Behind the Persistence of the Racial Wealth Gap?" Economic Commentary, February 28, Federal Reserve Bank of Cleveland. https://www.clevelandfed.org/newsroom-and-events/publications/economic-commentary/2019-economic-commentaries/ec-201903-what-is-behind-the-persistence-of-the-racial-wealth-gap.aspx.

Altonji, Joseph G., and Rebecca M. Blank. 1999. "Race and Gender in the Labor Market." In *Handbook of Labor Economics, Volume 3C*, edited by Orley C. Ashenfelter and David Card. Amsterdam: North-Holland.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen. 2019. "Who Becomes an Inventor in America? The Importance of Exposure to Innovation." *Quarterly Journal of Economics* 134, no. 2: 647–713.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94, no. 4: 991–1013.

Bhutta, Neil, Andrew C. Chang, Lisa J. Dettling, and Joanne W. Hsu. 2020. "Disparities in Wealth by Race and Ethnicity in the 2019 Survey of Consumer Finances." FEDS Notes. Washington: Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/econres/notes/feds-notes/disparities-in-wealth-by-race-and-ethnicity-in-the-2019-survey-of-consumer-finances-20200928.htm.

Black, Sandra E., Paul J. Devereux, Petter Lundborg, and Kaveh Majlesi. 2020. "Poor Little Rich Kids? The Role of Nature versus Nurture in Wealth and Other Economic Outcomes and Behaviours." *Review of Economic Studies* 87, no. 4: 1683–725.

Bostic, Raphael. 2020. "A Moral and Economic Imperative to End Racism." Federal Reserve Bank of Atlanta. https://www.atlantafed.org/about/atlantafed/officers/executive_office/bostic-raphael/message-from-the-president/2020/06/12/bostic-a-moral-and-economic-imperative-to-end-racism.

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.

Card, David, and Alan B. Krueger. 1992. "School Quality and Black-White Relative Earnings: A Direct Assessment." *Quarterly Journal of Economics* 107, no. 1: 151–200.

Chin, Aimee. 2005. "Long-Run Labor Market Effects of Japanese American Internment during World War II on Working-Age Male Internees." *Journal of Labor Economics* 23, no. 3: 491–526.

Cook, Lisa D. 2020. "Racism Impoverishes the Whole Economy." *New York Times*, November 18. https://www.nytimes.com/2020/11/18/business/racism-impoverishes-the-whole-economy.html.

Daly, Mary C. 2021. "Confronting Inequity: A Framework for Change." Federal Reserve Bank of San Francisco, February 4. https://www.frbsf.org/our-district/about/sf-fed-blog/confronting-inequity-framework-for-change/.

Daly, Mary C., Bart Hobijn, and Joseph H. Pedtke. 2017. "Disappointing Facts about the Black-White Wage Gap." Federal Reserve Bank of San Francisco, September 5. https://www.frbsf.org/economic-research/publications/economic-letter/2017/september/disappointing-facts-about-black-white-wage-gap/.

Daly, Mary C., Bart Hobijn, and Joseph H. Pedtke. 2020. "Labor Market Dynamics and Black-White Earnings Gaps." *Economics Letters* 186, no. 2: 108807.

Darity, William A., and Patrick L. Mason. 1998. "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender." *Journal of Economic Perspectives* 12, no. 2: 63–90.

Darity, William A., and Rhonda M. Williams. 1985. "Peddlers Forever? Culture, Competition, and Discrimination." *American Economic Review* 75, no. 2: 256–61.

Del Rio, Coral, and Olga Alonso-Villar. 2015. "The Evolution of Occupational Segregation in the United States, 1940–2010: Gains and Losses of Gender-Race/Ethnicity Groups." *Demography* 52, no. 3: 967–88.

Denton, Nancy A., and Douglas S. Massey. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, Mass.: Harvard University Press.

Dippel, Christian. 2014. "Forced Coexistence and Economic Development: Evidence from Native American Reservations." *Econometrica* 82 (November): 2131–65.

Espinosa, Lorelle L., Jonathan M. Turk, Morgan Taylor, and Hollie M. Chessman. 2019. *Race and Ethnicity in Higher Education: A Status Report*. Washington: American Council on Education. https://tacc.org/sites/default/files/documents/2019-03/race-and-ethnicity-in-higher-education.pdf.

García, Jorge Luis, James J. Heckman, Duncan Ermini Leaf, and María José Prados. 2020. "Quantifying the Life-Cycle Benefits of an Influential Early-Childhood Program." *Journal of Political Economy* 128, no. 7: 2502–41.

Herring, Cedric. 2009. "Does Diversity Pay? Race, Gender, and the Business Case for Diversity." *American Sociological Review* 74, no. 2: 208–24.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. 2019. "The Allocation of Talent and US Economic Growth." *Econometrica* 87, no. 5: 1439–74.

Kline, Patrick M., Evan K. Rose, and Christopher R. Walters. 2021. "Systemic Discrimination among Large U.S. Employers." Working Paper 29053. Cambridge,

Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29053.

Kochhar, Rakesh, and Anthony Cilluffo. 2018. "Income Inequality in the U.S. Is Rising Most Rapidly among Asians." Pew Research Center, July 12. https://www.pewresearch.org/social-trends/2018/07/12/income-inequality-in-the-u-s-is-rising-most-rapidly-among-asians/.

Mason, Patrick L., and Rhonda M. Williams. 1997. "Introduction: The Janus-Face of Race: Reflections on Economic Theory." In *Race, Markets, and Social Outcomes*, edited by Patrick L. Mason and Rhonda M. Williams. Boston: Springer.

Matthews, Madison, and Valerie Wilson. 2018. "Separate Is Still Unequal: How Patterns of Occupational Segregation Impact Pay for Black Women." Blog post, August 6, Economic Policy Institute. https://www.epi.org/blog/separate-is-still-unequal-how-patterns-of-occupational-segregation-impact-pay-for-black-women/.

Neumark, David. 2018. "Experimental Research on Labor Market Discrimination." *Journal of Economic Literature* 56, no. 3: 799–866.

Noel, Nick, Duwain Pinder, Shelley Stewart III, and Jason Wright. 2019. "The Economic Impact of Closing the Racial Wealth Gap. McKinsey and Company, August 13. https://www.mckinsey.com/industries/public-and-social-sector/our-insights/the-economic-impact-of-closing-the-racial-wealth-gap.

Oliver, Melvin L., and Thomas Shapiro. 1995. *Black Wealth/White Wealth: A New Perspective on Racial Inequality*. New York: Routledge.

O'Neill, June. 1990. "The Role of Human Capital in Earnings Differences between Black and White Men." *Journal of Economic Perspectives* 4, no. 4: 25–45.

Pager, Devah, and Hana Shepherd. 2008. "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets." *Annual Review of Sociology* 34:181–209.

Pager, Devah, Bruce Western, and Bart Bonikowski. 2009. "Discrimination in a Low-wage Labor Market: A Field Experiment." *American Sociological Review* 74, no. 5: 777–79.

Parks, Virginia. 2012. "The Uneven Geography of Racial and Ethnic Wage Inequality: Specifying Local Labor Market Effects." *Annals of the Association of American Geographers* 102, no. 3: 700–725.

Peri, Giovanni. 2014. "Do Immigrant Workers Depress the Wages of Native Workers?" IZA World of Labor. https://wol.iza.org/articles/do-immigrant-workers-depress-the-wages-of-native-workers/long.

Peterson, Dana M., and Catherine L. Mann. 2020. "Closing the Racial Inequality Gaps: The Economic Cost of Black Inequality in the U.S." Citi GPS: Global Perspectives and Solutions. https://www.citivelocity.com/citigps/closing-the-racial-inequality-gaps/.

Rodgers, William M. 2019. "Race in the Labor Market: The Role of Equal Employment Opportunity and Other Policies." *Russell Sage Foundation Journal of the Social Sciences* 5, no. 5: 198–220.

Rose, Stephen J. 2017. *Mismatch: How Many Workers with a Bachelor's Degree Are Overqualified for Their Jobs?* Washington: Urban Institute. https://www.urban.org/sites/default/files/publication/87951/college_mismatch_final.pdf.

Rothstein, Richard. 2017. *The Color of Law: A Forgotten History of How Our Government Segregated America*. New York: W. W. Norton.

Spriggs, William E., and Rhonda M. Williams. 2000. "What Do We Need to Explain about African American Unemployment?" In *Prosperity for All? The Economic Boom and African Americans*, edited by Robert Cherry and William M. Rodgers. New York: Russell Sage Foundation.

Stewart, Shelley, III, Michael Chui, James Manyika, J. P. Julien, Vivian Hunt, Bob Sternfels, Jonathan Woetzel, and Haiyang Zhang. 2021. "The Economic State of Black America: What Is and What Could Be. McKinsey and Company, June 17. https://www.mckinsey.com/featured-insights/diversity-and-inclusion/the-economic-state-of-black-america-what-is-and-what-could-be.

Treuhaft, Sarah, Justin Scoggins, and Jennifer Tran. 2014. "The Equity Solution: Racial Inclusion Is Key to Growing a Strong New Economy." Research Brief. Oakland, Calif.: Policy Link. https://www.policylink.org/sites/default/files/Equity_Solution_Brief.pdf.

Turner, Ani. 2018. *The Business Case for Racial Equity: A Strategy for Growth*. Battle Creek, Mich.: W. K. Kellogg Foundation. https://altarum.org/sites/default/files/uploaded-publication-files/WKKF%20Business%20Case%20for%20Racial%20Equity.pdf.

Williams, Jhacova, and Valerie Wilson. 2019. *Black Workers Endure Persistent Racial Disparities in Employment Outcomes*. Washington: Economic Policy Institute. https://www.epi.org/publication/labor-day-2019-racial-disparities-in-employment/.

Williams, Rhonda M., and William E. Spriggs. 1999. "How Does It Feel to Be Free? Reflections on Black-White Economic Inequality in the Era of 'Color-Blind' Law." *Review of Black Political Economy* 27, no. 1: 9–21.

Wilson, Valerie, and Janelle Jones. 2018. *Working Harder or Finding It Harder to Work: Demographic Trends in Annual Work Hours Show an Increasingly Fractured Workforce*. Washington: Economic Policy Institute. https://www.epi.org/publication/trends-in-work-hours-and-labor-market-disconnection/.

Wilson, Valerie, and William M. Rodgers. 2016. *Black-White Wage Gaps Expand with Rising Wage Inequality*. Washington: Economic Policy Institute. https://www.epi.org/publication/black-white-wage-gaps-expand-with-rising-wage-inequality/.

Comments and Discussion

COMMENT BY

NICOLE FORTIN Sixty years after the passage of the Civil Rights Act, we will remember 2020 not only as the start of the COVID-19 pandemic but also because of the waves of protest against anti-Black racial injustice that followed the death of George Floyd at the hands of the police. For many economists, it was not difficult to acknowledge the existence of racism in law enforcement and the justice system. Economists have long been aware of this type of anti-Black discrimination and studied its extent and origin (Rehavi and Starr 2014; Goncalves and Mello 2021).

However, in terms of discrimination in the labor market, many are doubtful because it does not make economic sense according to Becker's (1957) taste-based theory of discrimination. Firms leave money on the table when they discriminate against Black workers by not employing them. Nevertheless, numerous papers, starting with Bertrand and Mullainathan (2004) experimenting with correspondence resumes, have found that job applicants with Black-sounding names received significantly lower call-back rates than job applicants with white-sounding names. More recently, in a much larger experiment, Kline, Rose, and Walters (2021) found that company-specific racial gaps in contacts with applicants are negatively correlated with firm profitability: more profitable firms are less biased against Black applicants.[1] Reducing biases against minority groups may reap economic benefits.

Buckman, Choi, Daly, and Seitelman take this point further by asking how much aggregate output would be gained if differences in labor market

1. The authors find that the contact gaps are highly concentrated in particular companies: 23 out of 108 Fortune 500 US employers are responsible for nearly half of lost contacts to Black applicants in the experiment.

15985-02b_BuckmanCom&GD-4thPgs.indd 112 5/12/22 12:09 PM

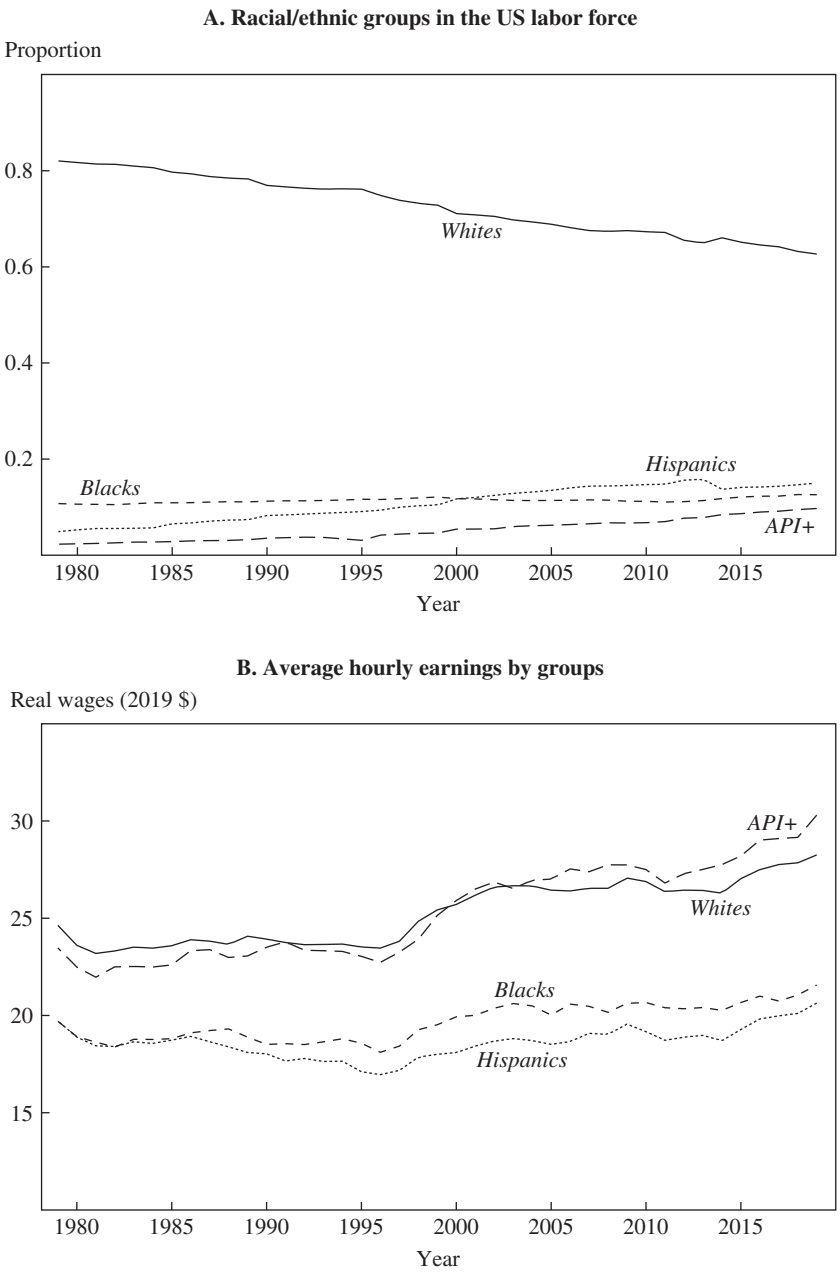
outcomes across racial/ethnic groups were erased. Now the argument that more equitable outcomes matter for both the level of GDP and the process of sustained economic growth will be even stronger in the future. Indeed, as America needs to welcome more immigrants from around the world, these will primarily fuel the ranks of minority groups. As shown in figure 1, panel A, the fraction of workers from the long-standing majority group, non-Hispanic white workers, is slowly diminishing over time: non-Hispanic white workers represented 63 percent of workers in 2019, down from 82 percent in 1979.[2]

More precisely, this ambitious, thought-provoking paper asks, "How much larger would the US economic pie be if opportunities and outcomes were more equally distributed by race and ethnicity?" The authors evaluate the impact on the labor contribution to GDP over thirty years in light of the question, What if Black and Hispanic workers had outcomes similar to those of non-Hispanic white workers? They focus on five outcomes of interest: employment, hours of work, educational attainment, educational utilization, and earnings gaps not explained by these and other productivity-related indicators. There are three notable features of this paper by comparison with the previous literature on the benefits of closing labor market gaps.

First, the authors consider a wider span of racial and ethnic groups that includes the typical non-Hispanic Black group, but also other groups with a high percentage of foreign-born workers, namely, Hispanic and other non-Hispanic, non-white, and non-Black individuals (e.g., Asian, Pacific Islander, American Indian) called API+. This addition reflects existing disparities in the labor market, as shown in figure 1, panel B. Hispanic workers have lower real average hourly earnings than Black workers, while API+ workers appear to have higher earnings than the majority group since 2000. The latter implies that in the construction of counterfactuals below, the substitution of non-Hispanic white earnings value for racial and ethnic groups will generally not apply to API+ groups. This grouping innovation opens up several issues regarding immigration policies and the potential (or lack thereof) to close differences in educational attainment for adult immigrants. These issues are discussed in more detail below.

2. These computations use the demographic groups defined by Buckman, Choi, Daly, and Seitelman: "White" for non-Hispanic white workers, "Black" for non-Hispanic Black workers, "Hispanic" for Hispanic workers of all races, and API+ for all remaining non-Hispanic, non-white, and non-Black individuals (e.g., Asian, Pacific Islander, American Indian). The analysis also focuses on employed workers age 25–64 using the CPS-MORG data, but uses the same variable construction as in Fortin, Lemieux, and Lloyd (2021).

Figure 1. Demographic Weights and Average Hourly Earnings by Racial/Ethnic Groups



Source: Author's calculations using CPS-MORG 1979-2019, employed workers age 25–64, https://www.nber.org/research/data/current-population-survey-cps-data-nber.

15985-02b_BuckmanCom&GD-4thPgs.indd 114 5/12/22 12:09 PM

Second, by performing the analysis annually for thirty years, from 1990 to 2019, the paper shows that the potential gains from erasing inequities are increasing over time. Unfortunately, starting in 1990 rather than 1980 misses a critical historical period in the erosion of policies and labor market institutions meant to support the most vulnerable workers, as shown in DiNardo, Fortin, and Lemieux (1996) and documented below. Therefore, the paper bypasses important insights to be learned from the past.

Third, the particular focus on employment opportunities helps connect the dots between the above well established but narrow experimental literature on discrimination in callback rates and more significant consequences for the labor market. Disparities in employment and hours worked are an often neglected dimension of this literature, and this paper fills a critical gap.

METHODS AND RESULTS The authors present two sets of counterfactuals; the first focuses on closing the gaps in earnings, the second on closing gaps in underlying factors to identify which factor yields the greatest gains. In the first instance, a group-specific counterfactual labor earnings contribution to total GDP is computed by multiplying the group's worker share and the average annual earnings of the majority group when group-specific earnings are lower. These counterfactual earnings, obtained by changing prices, provide the basis against which incremental gains from changing each factor will be assessed. The gains themselves are computed as the difference between observed group-specific labor earnings and the counterfactual group-specific labor earnings when minority group earnings are brought to at least the level of the majority group. For Black workers, the economic gains from getting counterfactual white earnings have remained relatively steady in percentage terms; their labor earnings contribution to GDP would have increased from 3.8 to 4.6 percent over the sample period. On the other hand, the overall counterfactual gains from labor income shows more substantial increases, from about 7.6 percent in 1990 to 10.2 percent in 2019, owing to the Hispanic share rising over the period.

Another set of counterfactuals is constructed at a disaggregated cell level (thirty-two cells for age, gender, and racial/ethnic group). The first two steps evaluate changes in employment rates and average weekly hours when the minorities' quantities are lower than the non-Hispanic white quantities, priced at the group-specific average hourly wage. In terms of contributions to labor earnings, employment rates stand at 2.1 percent and 2.2 percent in 1990 and 2019. The contributions of hours are much lower at 0.7 percent and 0.9 percent. Both contributions vary less over time.

In a third step, the counterfactuals adjust the entire vector of educational attainment in each cell to that of the majority in cases where the minority shares of those with at least some college are lower than in the majority group. From this imputed educational distribution, the authors derive new group-specific wages, changing prices in construction of this counterfactual. A fourth step boosts group-specific average earnings by changing the education utilization for racial and ethnic minorities when it is lower than for non-Hispanic white workers. Changing prices and quantities implies multiplicative effects or interaction terms, which lead to adding up issues, but they appear minor, as acknowledged by the authors. Here the percentage gains of labor earnings contributions from education and utilization show sizeable increases from 1.2 percent to 2.9 percent and from 0.1 percent to 0.6 percent, respectively, in 1990 and 2019. Finally, a fifth step adjusts whatever remains in residual earnings to match those of the majority group.

The main takeaway is that eliminating racial and ethnic disparities in employment rates and educational attainment makes the most significant contributions among the factors analyzed. However, when looking at the relative importance of these factors for Black workers, the authors find much smaller gains from equalizing education in 2019, in particular. More importantly, consistent with prior research, the authors find that residual earnings gaps play the largest role in generating GDP gains. These gains arising from unexplained sources are largest for Black workers and do not diminish over time. This finding is consistent with the cited literature that finds little progress in reducing residual earnings gaps over the sample period. Further disaggregation shows employment gaps and hours gaps play a more significant role for Black men than Black women, especially in 2019, while it is the reverse for educational attainment.

These results suggest the existence of barriers to equal educational and labor market opportunities for minority groups. However, it is also possible that policies meant to apply universally have differential impacts across racial/ethnic groups, making potential discriminatory effects more difficult to discern.

THE LABOR ECONOMIST'S VIEWPOINT The following points of discussion come from the viewpoint of the labor economist in terms of policies that might help materialize the gains. These policies include immigration policies, affirmative action measures, and labor market institutions, namely, labor unions and minimum wages. These two last sets of measures long protected the most vulnerable workers, including minorities. Indeed, Farber and others (2021) find that from the 1940s to the 1990s, unions conferred a sizeable (10–20 percent) family income premia to non-white and less

educated households, who were overrepresented among union members.[3] Using compelling research designs, Derenoncourt and Montialoux (2021) show that the 1967 extension of the minimum wage accounts for more than 20 percent of the reduction in the racial earnings and income gap during the civil rights era. However, these institutions have been eroded since the 1980s (DiNardo, Fortin, and Lemieux 1996), a decade earlier than the facts documented in this paper. The stylized facts presented below will suggest a potential role for the erosion of institutions in the lack of improvement in the labor market outcomes of minority groups, especially Black workers. However, establishing a clear link between the weakening of these policies and institutions and the lack of progress of minorities would require a more thorough and rigorous investigation, as demonstrated in Farber and others (2021) and Derenoncourt and Montialoux (2021).

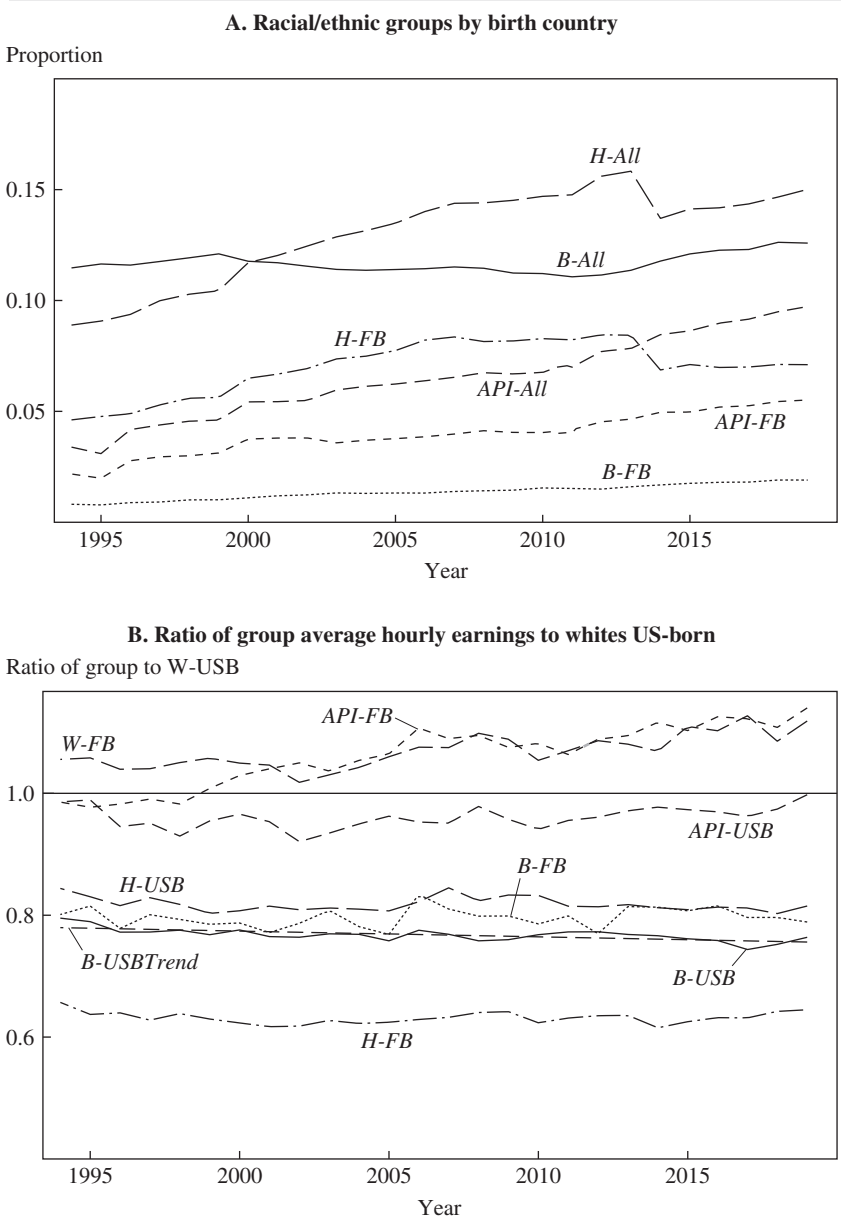
The first important point is that the definition of ethnic and racial groups matters. Using the same four groups as the authors, as shown in figure 2, panel B, the earnings of Black workers surpass those of Hispanic workers, and the earnings of the API+ group surpass those of white workers in 2000. However, equity issues are different when considering adult immigrant workers who arrive in the United States after completing their education. For example, many Hispanic "economic refugees" from Mexico have a much lower level of education than US-born Hispanics (Chiquiar and Hanson 2005). While there are state-level education policies directed at the children of immigrants, who would be first-generation college-goers, it is unfeasible to upgrade the educational attainment of foreign-born Hispanics to the level suggested by the exercise above. Conversely, immigrants from large Asian countries, such as China and India, are more likely positively selected, partly because of the per country ceilings on permanent immigration visas and other visa issues.[4] On the other hand, some US-born members of the API+ group, such as US-born Pacific Islanders and American Indians, may face some persistent labor market disadvantages. As acknowledged by the authors, the API+ group is quite heterogeneous.

3. Farber and others (2021) point out that in this era, part of the overrepresentation of Black workers in unions arose merely as a by-product of unions organizing lower-skilled areas of the economy, which were disproportionately non-white.

4. To ensure that certain nationalities do not dominate immigration flows to the United States, permanent immigrants (family-based and employment-based combined) from a single country cannot exceed 7 percent of the total number of people immigrating in a single year. Budiman (2020) documents that in 2018, after Mexico, which accounted for 25 percent of the US immigrant population, the next two largest origin countries were China and India at 6 percent each. See also Maynard and Seeborg (2014), Liao and Seeborg (2015), and Orrenius, Zavodny, and Kerr (2012) for the impact of special visas.

15985-02b_BuckmanCom&GD-4thPgs.indd 117 5/12/22 12:09 PM

Figure 2. Differences by Country of Birth Status among Racial/Ethnic Groups



Source: Author's calculations using CPS-MORG 1979-2019, employed workers age 25–64, https://www.nber.org/research/data/current-population-survey-cps-data-nber.

Note: W—white workers; B—Black workers; H—Hispanic workers; FB—foreign-born; USB—US-born. Birth country data available from 1994.

15985-02b_BuckmanCom&GD-4thPgs.indd 118 5/12/22 12:09 PM

The share of foreign-born workers among minority groups is not trivial. Figure 2, panel A, displays the trends in the demographic weights of all employed workers from each minority group, along with the demographic weights of foreign-born in each group.[5] The ratio of the two group-specific trends gives the share of foreign-born within each group (not displayed). For example, we can compute that in 2010, the share of foreign-born was 56 percent among Hispanic workers, more than 60 percent among API+ workers; it is still substantial at 14 percent among Black workers but was much lower at 4 percent for non-Hispanic white workers.[6] Immigration policies play a significant role in the racial/ethnic composition of the US labor force, thus in its contribution to GDP.

Differences of average hourly earnings in the ranking of groups between the US-born and the foreign-born, displayed in figure 2, panel B, are stunning. Among the US-born, Hispanic workers do better than Black workers, and API+ workers do not do as well as non-Hispanic white workers. Among the foreign-born, Black workers do better than Hispanic workers, and API+ workers are similar to non-Hispanic white workers. Thus, the wage advantage of the API+ group mostly comes from their immigrant status. Therefore, the decision made by the authors to use non-Hispanic white wages as the counterfactual of choice is empirically well founded, given the relatively lower share of positively selected foreign-born among them, although that share is growing over time. However, when we distinguish the US-born and foreign-born, we find that relative to their US-born non-Hispanic white counterparts the wages of US-born Black workers are declining over time, a grimmer situation than usually found in the literature. What might be the sources of this undesirable trend?

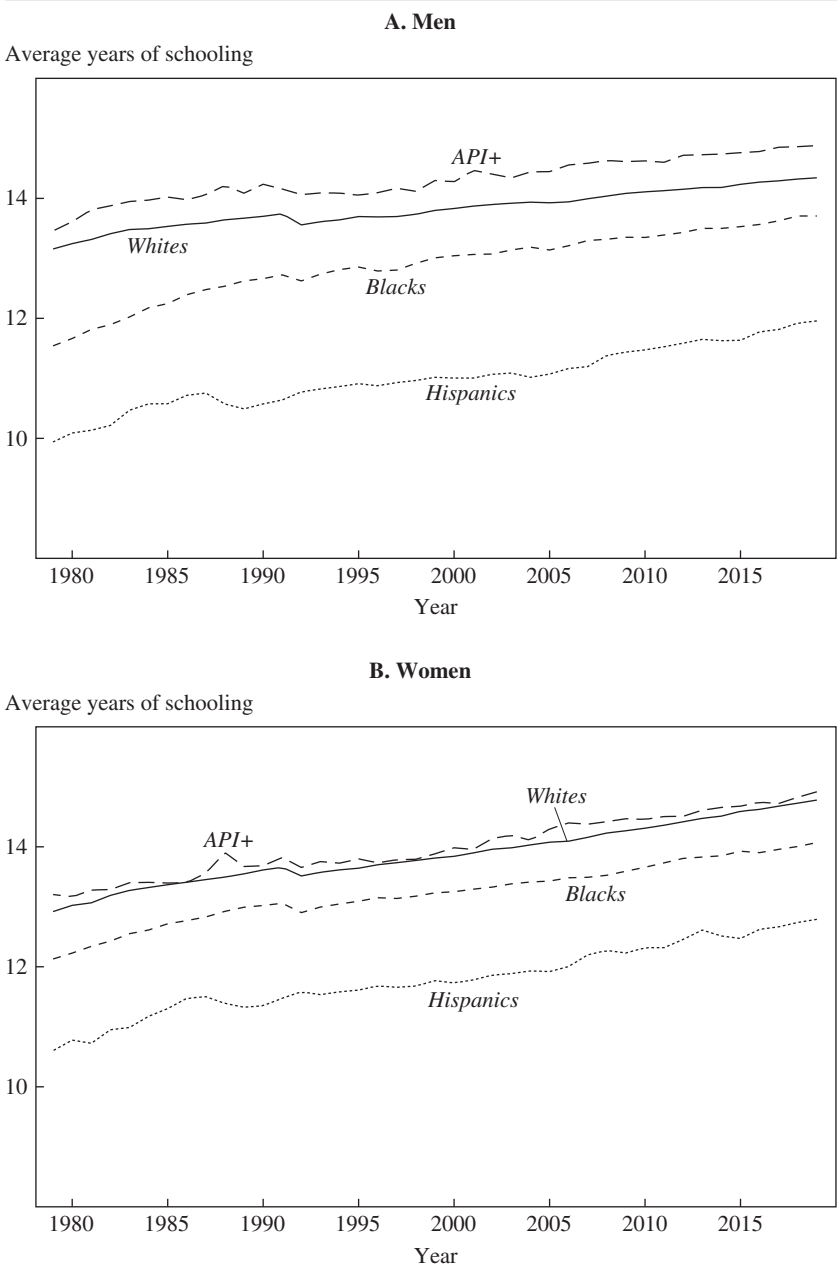
The authors have already shown the striking differences in educational attainment between Black workers and non-Hispanic white workers, in particular, and that these gaps are not converging. For instance, figure 3 goes back a decade earlier in time and shows that the improvement in the annual growth in years of schooling among Black workers slowed down by

5. For example, the B-All line shows the evolution of the demographic weight of all Black workers, ranging from 11 to 12.6 percent of workers, while the B-FB line traces the evolution of the demographic weight of Black workers who are foreign-born, which ranges from 0.8 to 1.9 percent of workers. As a result, the percentages of foreign-born among Black workers range from 6.6 to 14.9 percent. For each of the other minority groups, the evolution of the demographic weights of foreign-born members (H-FB and API-FB) are traced with less solid lines.

6. For clarity, the evolution of the foreign-born among non-Hispanic white workers is not displayed; we can compute that it had grown from 4 percent in 2010 to 7.5 percent in 2019.

15985-02b_BuckmanCom&GD-4thPgs.indd 119 5/12/22 12:09 PM

Figure 3. Educational Attainment by Racial/Ethnic Groups



Source: Author's calculations using CPS-MORG 1979-2019, employed workers age 25–64, https://www.nber.org/research/data/current-population-survey-cps-data-nber.

15985-02b_BuckmanCom&GD-4thPgs.indd 120 5/12/22 12:09 PM

half after 1996.[7] That year, bans on affirmation action in higher education, also characterized as a move toward race-neutral policies, began to take hold in several large states (Howell 2010).[8]

Affirmative action policies, which began in the late 1960s, allowed admissions departments at selective higher education institutions to consider race as a factor when admitting new students. Subsequently perceived as reverse discrimination, these policies were subjected to legal challenges. Howell (2010) studies the replacement policies instituted in state public colleges in California, Florida, and Texas. These policies favored high-performing (e.g., top 10 percent) students from economically disadvantaged high schools and were thought to provide a practical alternative. However, Howell (2010) argues that these programs would not be successful at restoring minority representation at the most selective colleges, given that affirmative action was helping a wider range of minority students and not only the highest performers.

Labor unions are institutions that historically supported workers from all groups. Union jobs were those "good jobs" that allowed workers to access the middle class. However, since the mid-twentieth-century union peak (Farber and others 2021), there has been a significant decline in unionization rates across many industries and states.[9] More recently, this decline was exacerbated by right-to-work (RTW) laws, which prevent unions from collecting union dues from nonmembers and thus generate "free-rider" problems (Ichniowski and Zax 1991). Initially adopted in the late 1940s, RTW laws have made a comeback in several states since 2001 (Fortin, Lemieux, and Lloyd 2021).[10] At the federal level, a 2019 Supreme Court ruling has extended RTW to all public sector employees—local, state, and federal—within the United States.

Farber and others (2021) documented that at their mid-twentieth-century peak, unions were organizing groups that were disproportionately nonwhite. However, it is less well known how much larger union coverage

7. Before 1996, years of completed education among Black men grew by 0.08 a year on average, from 11.5 in 1979 to 12.9 in 1995. That rate, at 0.04, was half from 1996 to 2019, when the average reached 13.7 years.

8. These include, by year of adoption, California (1996), Texas (1996–2003), Washington (1998), Florida (1999), Michigan (2006), Nebraska (2008), Arizona (2010), New Hampshire (2012), Oklahoma (2012), and Idaho (2020).

9. The 1980s decline in unionization rates was initially propelled by the deregulation of federally regulated industries in the Reagan area.

10. These states include, by year of RTW adoption, Oklahoma (2001), Indiana (2012), Michigan (2013), Wisconsin (2015), West Virginia (2016), and Kentucky (2017). On the other hand, a handful of states have turned down RTW laws or adopted bans: Missouri (2018), New Mexico (2019), and New Hampshire (2017 and 2021).

15985-02b_BuckmanCom&GD-4thPgs.indd 121 5/12/22 12:09 PM

rates were among Black workers than among other groups and therefore, how much steeper the decline was in union coverage for Black workers, both men and women.

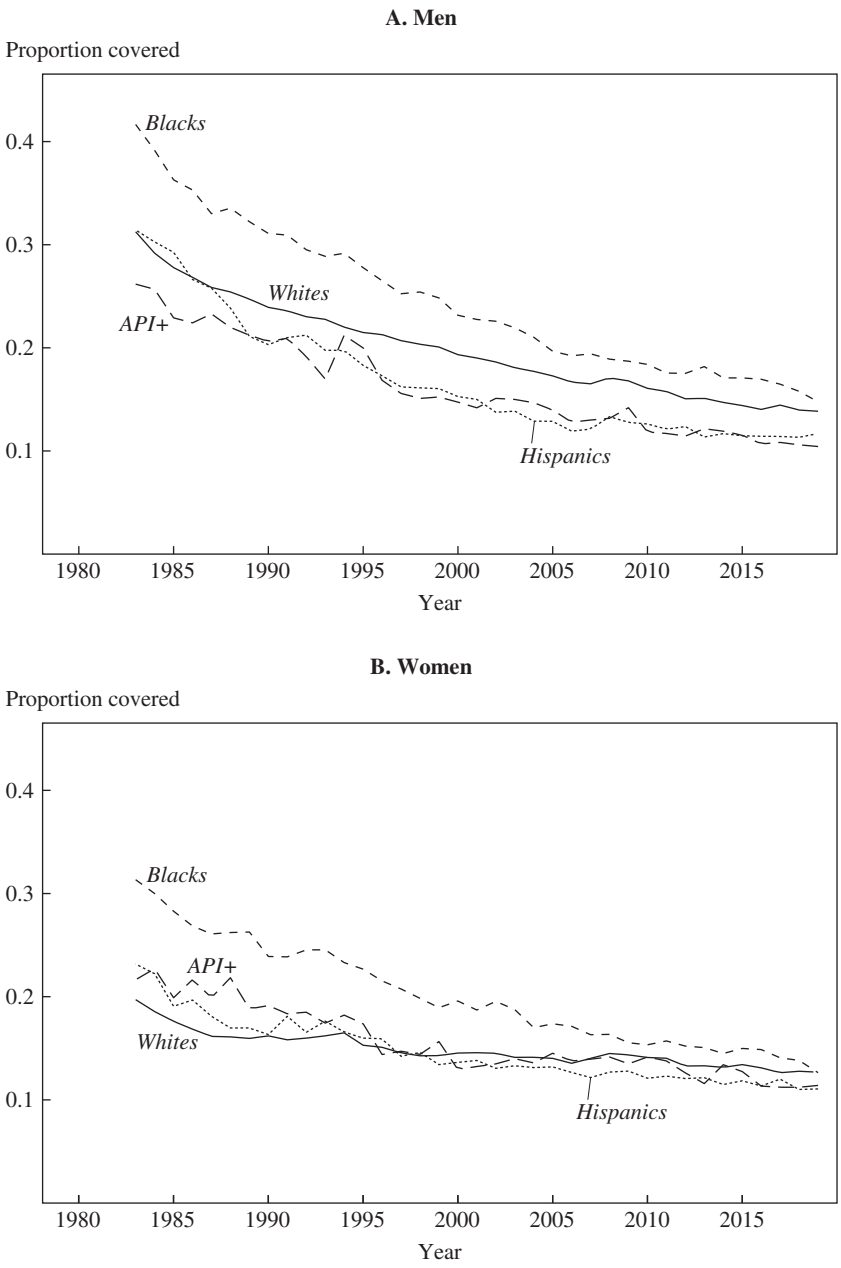
Figure 4 illustrates that steeper decline, especially for Black men: union coverage rates dropped from 42 percent to 15 percent between 1983 and 2019, while they dropped from 31 percent to 12 percent for Black women. Not surprisingly, a dramatic decline in the power of unions to extract rents and boost the union/nonunion wage gap accompanied the decline in union density. Figure 5 displays the level of average hourly earnings by union status (panel A) and the union premium computed as the ratio of average hourly earnings of unionized to nonunion workers minus one (panel B). Again, the trends are displayed for non-Hispanic white workers and the groups whose earnings are consistently lower than those of the majority group. A few striking stylized facts emerge from this graph.

In panel A, the white advantage persists even over unionized minority workers, although before 1995 nonunion white workers do barely better than unionized Hispanic or Black workers. The gap between these last two groups is almost indistinguishable, showing that unions could equalize wages across minority groups. The fact that nonunion wages converge to union wages among non-Hispanic white workers illustrates one reason for disaffection with unions among the majority group. Panel B illustrates this fact in terms of the erosion of the union/nonunion premium for this group. It also shows the sharper decline in the union premium for Black workers. Black workers in the middle of the wage distribution were facing strong headwinds to get ahead. The good unionized manufacturing jobs that gave them access to the middle class were disappearing (Autor 2010).

Meanwhile, the proportion of minimum wage workers was consistently higher among Black workers than among non-Hispanic white workers, although it was much smaller than among Hispanic workers, as shown in figure 6. In particular, after the Great Recession of 2008, 7 percent of Black men were minimum wage workers, while that proportion was 3 percent among non-Hispanic white men and 11 percent among Hispanic men. Among women over the last decade, the proportion of minimum wage workers was highest among Hispanic workers at 18 percent, triple that of non-Hispanic white workers at 6 percent. However, the proportion of female minimum wage workers was similar among Black workers, 9.6 percent, and API+ workers, 9 percent, reflecting that this last group is substantially heterogeneous, as discussed above. This heterogeneity also explains why, when the authors perform their analysis by gender, they find different results for API+ men and women.

15985-02b_BuckmanCom&GD-4thPgs.indd 122 5/12/22 12:09 PM

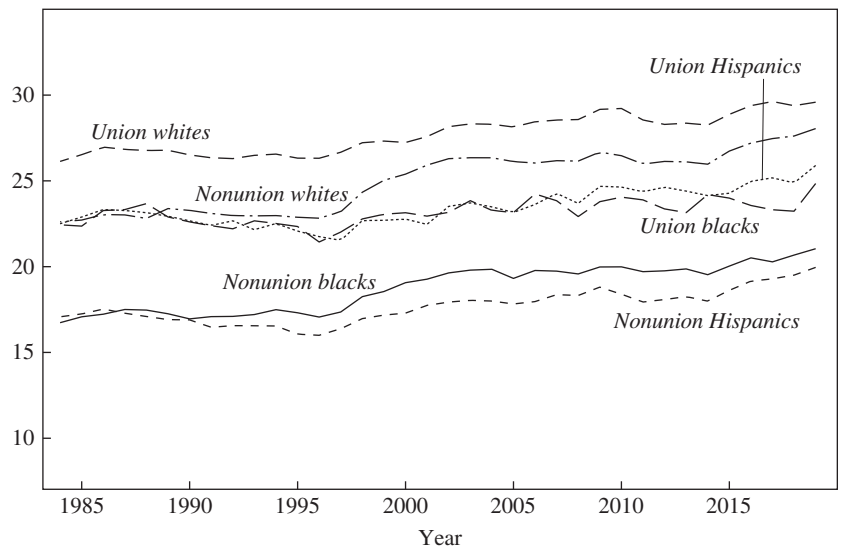
Figure 4. Union Coverage by Racial/Ethnic Groups



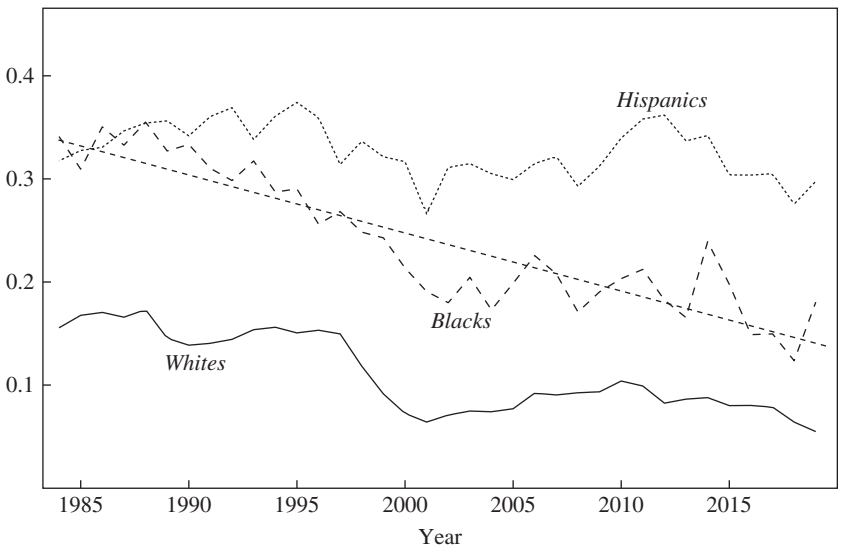
Source: Author's calculations using CPS-MORG 1979-2019, employed workers age 25–64, https://www.nber.org/research/data/current-population-survey-cps-data-nber.

Figure 5. Union Premium across Racial/Ethnic Groups

A. Average hourly earnings by union status and racial/ethnic groups

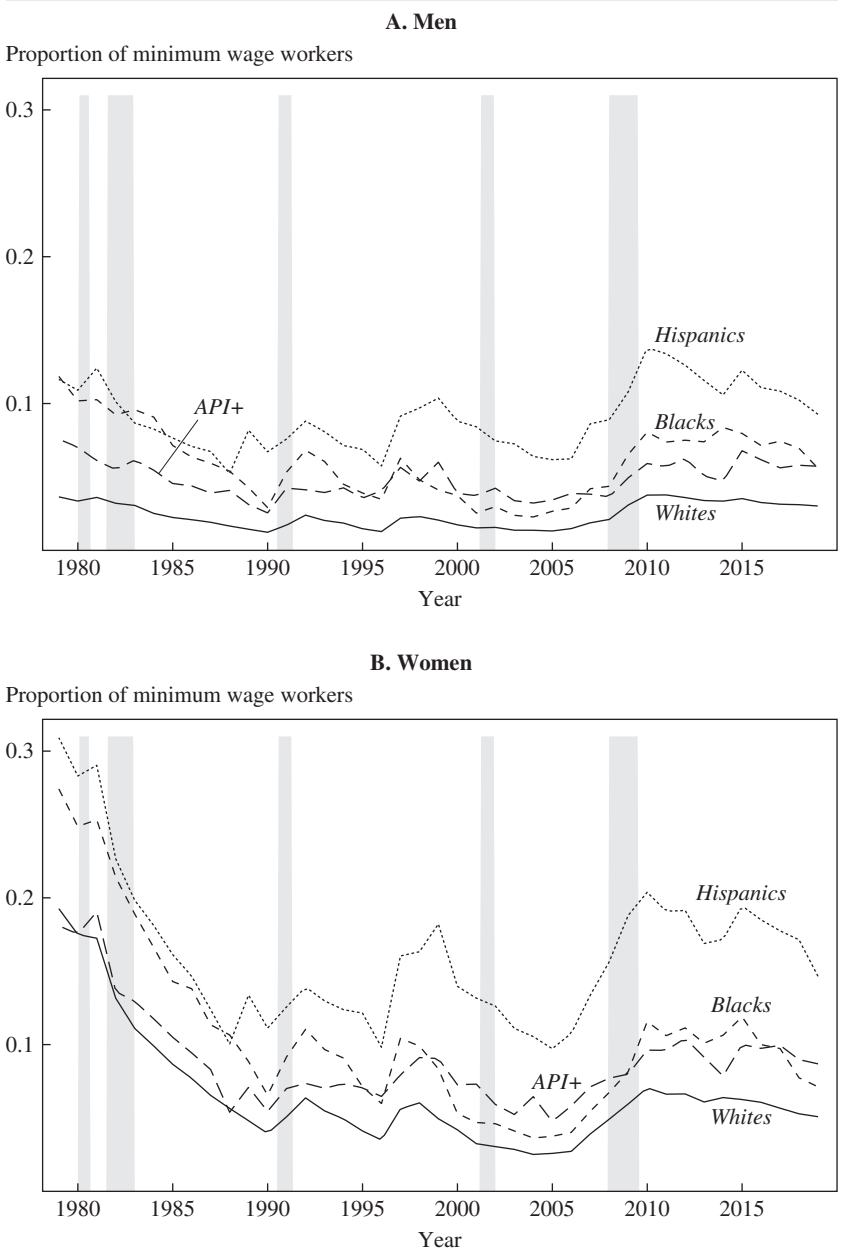


B. Union/nonunion premium by racial/ethnic group



Source: Author's calculations using CPS-MORG 1979-2019, employed workers age 25–64, https://www.nber.org/research/data/current-population-survey-cps-data-nber.

Figure 6. Proportion of Minimum Wage Workers across Racial/Ethnic Groups



Source: Author's calculations using CPS-MORG 1979-2019, employed workers age 25–64, https://www.nber.org/research/data/current-population-survey-cps-data-nber.

Note: Within 15 percent of effective minimum wage. Shaded areas indicate recessions.

Unlike the case of unions, minimum wages have remained a popular labor market institution. Indeed, in 2012, a new wave of proponents surged under the banner of "Fight for $15," when two hundred fast-food workers walked off the job to demand $15/hour and union rights in New York City.[11] By 2018, several cities (San Francisco, Berkeley, Los Angeles, and New York City) had adopted the prescription, which was almost triple the state minimum wage in other areas (e.g., in Georgia and Wyoming, it is $5.15/hour, noting employees subject to the Fair Labor Standards Act must be paid the federal minimum wage of $7.25/hour).

The movement coupled with the stagnation of federal minimum wages marks a new era of increasing minimum wage inequality. Figure 7, panel A, illustrates the divergence between state and federal minimum wages. Minimum wage workers in some states can earn almost twice as much as their counterparts in other states. This increasing disparity begs the question: Which racial/ethnic groups benefit or suffer from this trend? In 2019, the federal minimum wage of $7.25/hour prevailed in twenty US states, including eleven states in the South (Alabama, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, and Virginia) where 43 percent of Black workers live, in comparison to 27 percent of Hispanic workers. Thus, it will not come as a surprise that Black minimum wage workers are overrepresented in the bottom quartile of minimum wage workers, as shown in figure 7, panel B.[12] Although Hispanic workers are more likely than Black workers to be minimum wage workers, as shown in figure 6, panel B, they are less likely to be in the bottom quartile of minimum wages, given that many work in California, a higher minimum wage state.[13]

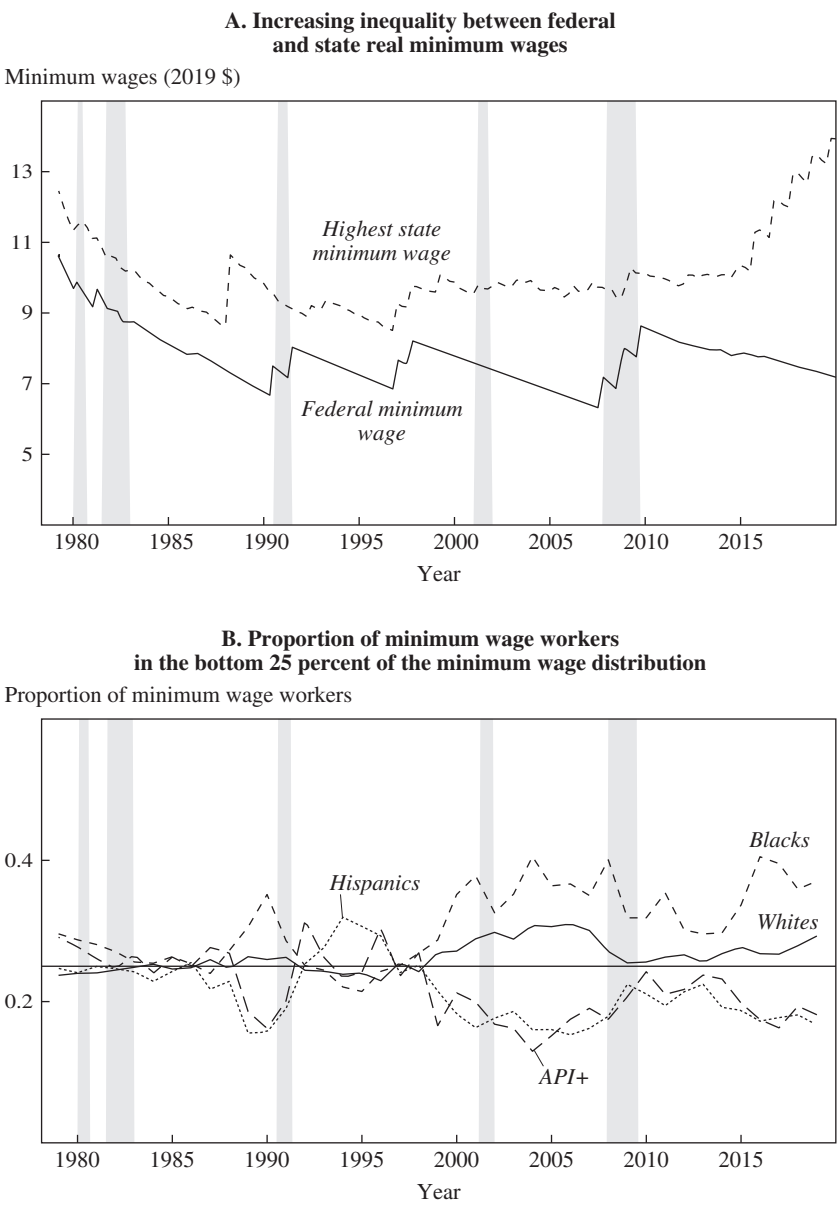
Following the recovery after the Great Recession, by comparison with non-Hispanic white workers, Black workers are twice as likely to be minimum wage workers and 20–50 percent more likely to be in the bottom quartile of minimum wages. Black workers are still more likely to be unionized than non-Hispanic white workers. However, their average unionized earnings are lower than those of the majority group, and the union premium they face has been declining over time. In sum, the labor market institutions that historically helped support Black workers are now less effective.

11. Fight for $15, https://fightfor15.org/.

12. Equal representation in the bottom quartile of minimum wages would be at 25 percent, that is, on the horizontal line. It corresponds to the share of non-Hispanic white minimum wage workers from 2009 to 2013 found in this bottom quartile.

13. In 2019, 27 percent of Hispanic workers lived in California by comparison with 6 percent of Black workers; see also Krogstad and Noe-Bustamante (2021).

Figure 7. Increasing Inequality within Minimum Wage Workers across Racial/Ethnic Groups



Sources: Author's calculations using CPS-MORG 1979-2019, employed workers age 25–64, https://www.nber.org/research/data/current-population-survey-cps-data-nber; minimum wage data from FRED and Kavya Vaghul and Ben Zipperer, "Historical State and Sub-state Minimum Wages," version 1.2.0, https://github.com/benzipperer/historicalminwage/releases/tag/v1.2.0.

Note: Within 15 percent of effective minimum wage. Shaded areas indicate recessions.

However, this applies less to Hispanic workers. They are less likely to be unionized than non-Hispanic white workers, but their average unionized earnings are comparable to Black workers, and the union premium they face has not declined as much over time. The average earnings of US-born Hispanic workers exceed that of Black workers, while the American dream of foreign-born Hispanic workers is still distant and may await the next generation. Indeed, leveling the quality of education across racial/ethnic neighborhoods and improving immigration policies are important but challenging objectives to achieve.

Given all the obstacles documented above, it is remarkable that Black earnings have merely stagnated, and it is not surprising that in the authors' analysis, the most significant gains to GDP for residual earnings equalization come from this group. While general disaffection with unions implies that they are not coming back (Farber 1990), could further erosion of union power be averted by reversing or banning RTW laws? The experience of the few states which did so has yet to be studied. On the minimum wage front, incremental increases to federal or state minimum wages would likely be a more feasible alternative than the push to $15, which in some cases might double prevailing minimum wages. Arguably, minimum wages that are too low may be an obstacle to innovation.[14] They fail to encourage firms to adopt more efficient technologies when earned income tax credit or similar income support programs implicitly boost low-skilled workers' reservation wages.

Finally, is there a role for direct antidiscrimination measures, such as the affirmative action policies discussed above? Nearly forty years ago, there was much optimism about these policies. Freeman and others (1973, 119) argued that "much of the improvement in the black economic position that took place in the late sixties appears to be the result of governmental and related anti-discriminatory activity associated with the 1964 Civil Rights Act." Unfortunately, as documented above, affirmative action measures lost favor from the 1990s onward.[15]

14. Lordan and Neumark (2018) find that minimum wage increases will give incentives for firms to adopt new technologies that replace workers. The flip side of job losses caused by automation following a minimum wage increase is that low minimum wages decrease firms' incentives to automate low-skilled jobs.

15. McCrary (2007) finds that court-ordered racial hiring quotas imposed on municipal police departments starting in 1973 led to a 14 percentage point gain in the fraction of African Americans among newly hired officers. However, he argues that by current standards, these measures were remarkably interventionist and among the most aggressive affirmation actions ever implemented in the United States.

15985-02b_BuckmanCom&GD-4thPgs.indd 128 5/12/22 12:09 PM

Today, the authors offer a more targeted but renewed optimism in "a new mindset—a mindset that sees large gaps by race and ethnicity as inefficient and a sign of misallocation, rather than an unfortunate, but efficient, outcome of a well-functioning market." In the twenty-first century, human interactions among young people, in particular, are changing. With the implied renewed mindsets, this optimism is hopefully well placed.

REFERENCES FOR THE FORTIN COMMENT

Autor, David. 2010. *The Polarization of Job Opportunities in the US Labor Market: Implications for Employment and Earnings*. Washington: Brookings Institution, Hamilton Project, and Center for American Progress.

Becker, Gary. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94, no. 4: 991–1013.

Budiman, Abby. 2020. "Key Findings about U.S. Immigrants." Pew Research Center, August 20. https://www.pewresearch.org/fact-tank/2020/08/20/key-findings-about-u-s-immigrants/.

Chiquiar, Daniel, and Gordon H. Hanson. 2005. "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States." *Journal of Political Economy* 113, no. 2: 239–81.

Derenoncourt, Ellora, and Claire Montialoux. 2021. "Minimum Wages and Racial Inequality." *Quarterly Journal of Economics* 136, no. 1: 169–228.

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

Farber, Henry S. 1990. "The Decline of Unionization in the United States: What Can Be Learned from Recent Experience?" *Journal of Labor Economics* 8, no. 1, pt. 2: S75–S105.

Farber, Henry S., Daniel Herbst, Ilyana Kuziemko, and Suresh Naidu. 2021. "Unions and Inequality over the Twentieth Century: New Evidence from Survey Data." *Quarterly Journal of Economics* 136, no. 3: 1325–85.

Fortin, Nicole M., Thomas Lemieux, and Neil Lloyd. 2021. "Labor Market Institutions and the Distribution of Wages: The Role of Spillover Effects." *Journal of Labor Economics* 39, no. S2: S369–S412.

Freeman, Richard B., R. A. Gordon, Duran Bell, and Robert E. Hall. 1973. "Changes in the Labor Market for Black Americans, 1948–72." *Brookings Papers on Economic Activity*, no. 1: 67–131.

Goncalves, Felipe, and Steven Mello. 2021. "A Few Bad Apples? Racial Bias in Policing." *American Economic Review* 111, no. 5: 1406–41.

Howell, Jessica S. 2010. "Assessing the Impact of Eliminating Affirmative Action in Higher Education." *Journal of Labor Economics* 28, no. 1: 113–66.

Ichniowski, Casey, and Jeffrey S. Zax. 1991. "Right-to-Work Laws, Free Riders, and Unionization in the Local Public Sector." *Journal of Labor Economics* 9, no. 3: 255–75.

Kline, Patrick M., Evan K. Rose, and Christopher R. Walters. 2021. "Systemic Discrimination among Large US Employers." Working Paper 29053. Cambridge, Mass.: National Bureau of Economic Research.

Krogstad, Manuel, and Luis Noe-Bustamante. 2021. "Key Facts about U.S. Latinos for National Hispanic Heritage Month." Pew Research Center, September 9. https://www.pewresearch.org/fact-tank/2021/09/09/key-facts-about-u-s-latinos-for-national-hispanic-heritage-month/.

Liao, Lu, and Michael Seeborg. 2015. "An Analysis of Earnings Differentials between College-Educated Chinese Immigrants and US Natives: Who Has the Advantage?" *Journal of Economic Insight* 41, no. 2: 1–18.

Lordan, Grace, and David Neumark. 2018. "People versus Machines: The Impact of Minimum Wages on Automatable Jobs." *Labour Economics* 52, no. 3: 40–53.

Maynard, Paige, and Michael Seeborg. 2014. "The Effect of Occupational Choice on the Earnings of College Educated Immigrants from China, India, the Philippines and Mexico." *Journal of Economic Insight* 40, no. 2: 19–35.

McCrary, Justin. 2007. "The Effect of Court-Ordered Hiring Quotas on the Composition and Quality of Police." *American Economic Review* 97, no. 1: 318–53.

Orrenius, Pia, Madeline Zavodny, and Emily Kerr. 2012. "Chinese Immigrants in the US Labor Market: Effects of Post-Tiananmen Immigration Policy." *International Migration Review* 46, no. 2: 456–82.

Rehavi, M. Marit, and Sonja B. Starr. 2014. "Racial Disparity in Federal Criminal Sentences." *Journal of Political Economy* 122, no. 6: 1320–54.

COMMENT BY

ERIK HURST What are the aggregate gains from creating equal opportunities across racial and ethnic groups in the United States? That is the question this paper tackles. This is an extremely difficult question to answer. The answer depends both on what is causing labor market differences across groups and on how a reference group can be defined to perform various counterfactuals.

A SIMPLE ILLUSTRATIVE EXAMPLE To set the stage, consider two groups, A and B, with members of each group comprising half of the total population. Suppose further that members of group A in a given period earn, on average, $75,000 a year while members in group B earn, on average, $25,000 per year in that period. What are the gains to this fictional economy of equating the average incomes of individuals in groups A and B?

One counterfactual would be to do the following. First, we could assume that there are some labor market barriers in this economy causing group B

to earn $50,000 less than they are currently earning. Specifically, one could assume that current average earnings of group A are the correct counterfactual as to what an undistorted economy for group B would look like. Further assuming that there are no real resource costs associated with equating incomes across groups, this counterfactual would imply that this fictitious economy could increase aggregate production by 50 percent if it was able to equate incomes across groups.

Another counterfactual is that group A is gaining due to labor market frictions at the expense of group B. Suppose, then, that in this counterfactual removing the labor market barriers increases the average income of group B members to $50,000 per year but simultaneously lowers the average income of group A members to $50,000 per year. In this case, equality is achieved but at a level of average income lower than what group A is currently earning. Again assuming that there are no real resource costs associated with equating incomes across groups, this counterfactual finds that there would be no aggregate gains associated with equating incomes across groups. The gains associated with improving the incomes of group B would come at the expense of reducing the incomes of group A. Assuming both groups are the same suggests that relative incomes should be the same, but more assumptions are needed to pin down the level of income.

The assumption of no real resource costs associated with equating incomes across groups is also important. Suppose we refer to group A as college-educated and group B as high school–educated. In this case, expenditures would potentially be needed to provide group B members with the additional human capital to make them as productive in the labor market as group A. A full counterfactual of the benefits to the aggregate economy from equating outcomes needs to take a stance on such resource costs.

In this paper, the goal is to think about the groups in terms of race or ethnicity. Group A is white individuals, with the other groups being Black, Hispanic, and other (including Asian and Pacific Islander). The paper assumes—and I concur—that in a world with no current or past barriers, the labor market outcomes across racial and ethnic groups should be similar. However, current and past discrimination almost certainly has resulted in different groups having different levels of human capital. As a result, it may cost large amounts of aggregate resources to equate labor market outcomes across groups. Additionally, it is likely that in this context group A has benefited from current and past discrimination toward the other groups. This implies the current average income of white individuals may not be the correct counterfactual for equating labor market opportunities across racial and ethnic groups. It is hard, therefore, to compute the aggregate

Table 1. 2017–2019 Raw Data

Group	*Share of sample*	*Average ACS labor income (incl. zeros)*	*Group-specific GDP (in trillions)*
Non-Hispanic white	0.62	$46,426	$4.9
Non-Hispanic Black	0.12	$29,163	$0.6
Hispanic	0.17	$28,891	$0.8
Other	0.09	$47,103	$0.7
Total ACS labor income			$7.0

Sources: IPUMS USA and the American Community Survey (ACS).

Note: The sample includes all individuals age 25–64 (inclusive). All dollar values are in current dollars. The last column measures the total labor income in each group as measured by the ACS during this period and is computed as the multiple of column 2, column 3, and 170 million individuals, with the latter number being the approximate BLS estimate of the aggregate population of individuals age 25–64 during this time period.

gains to equating labor market outcomes across groups without taking a stand on both the appropriate benchmark and the cost of equating incomes across groups.

The analysis in this paper follows the first potential counterfactual I highlighted above. The paper just asks the simple question of what would aggregate labor market income be if Black and Hispanic individuals earned the same average income as whites. To illustrate the key findings of the paper and to provide an additional counterfactual, I will use my own data from the pooled 2017–2019 American Community Survey (ACS) and the 1990 US Census. I downloaded these data directly from the IPUMS USA data set.[1] I restricted the sample to include all individuals (regardless of group quarters status) between the age of 25 and 64 (inclusive). I prefer the census and ACS data because they try to capture individuals residing in group quarters. As a result, my analysis differs very slightly from the findings in the paper using the Current Population Survey.

2017–2019 COUNTERFACTUALS Table 1 shows the raw data on share of the population and average income for different racial and ethnic groups from the pooled 2017–2019 ACS (columns 2 and 3). I define the racial groups analogously to how they are defined in the paper (with the "Other" group including primarily Asians and Pacific Islanders). The total size of the US population age 25–64 during this period, according to the US Bureau of Labor Statistics (BLS), was roughly 170 million individuals.[2]

1. IPUMS USA, "U.S. Census Data for Social, Economic, and Health Research," https://usa.ipums.org/usa/.

2. See US Census Bureau, "2019 Population Estimates by Age, Sex, Race, and Hispanic Origin," https://www.census.gov/newsroom/press-kits/2020/population-estimates-detailed.html.

15985-02b_BuckmanCom&GD-4thPgs.indd 132 5/12/22 12:09 PM

Table 2. 2017–2019 Counterfactual

Group	*Share of sample*	*Average ACS labor income (incl. zeros)*	*Group-specific GDP (in trillions)*
Non-Hispanic white	0.62	$46,426	$4.9
Non-Hispanic Black	0.12	$46,426	$0.9
Hispanic	0.17	$46,426	$1.3
Other	0.09	$47,103	$0.7
Total ACS labor income			$7.9

Source: Author's calculations.

Note: Table shows the simple counterfactual performed in the paper by equating the average labor incomes of Black and Hispanic individuals to that of white individuals.

Column 3 reports the average income earned last year (inclusive of zeros) for all individuals within each of the four racial and ethnic groups. All income is in current dollars. The last column is the multiple of the second column times the third column times 170 million for each group. The last column therefore measures the total aggregate labor income captured within the ACS data for each group. According to the ACS data, total labor income for 25- to 64-year-olds during this time period is roughly $7 trillion.

The main counterfactual in the paper is to assume that Black and Hispanic individuals earn the same average income as whites. Table 2 shows this main counterfactual using the ACS data set. Again, my numbers differ slightly than those in the paper because I am using the ACS data as opposed to the CPS. In particular, under this counterfactual, aggregate labor income (as measured by the ACS) would increase by about $0.9 trillion if Black and Hispanic individuals earned the same average income as white individuals. This represents a 12 percent gain in aggregate labor income.

1990 COUNTERFACTUAL Table 3 shows the raw data on share of the population and average income for different racial and ethnic groups from the 1990 US Census. This table has the same sample restrictions and layout as table 1 discussed above. According to the BLS, the total US population of 25- to 64-year-olds in 1990 was roughly 128 million. As seen from the table, total labor income measured in the US Census for this age group during this time period is roughly $4.6 trillion.

Notice that, relative to table 1, table 3 shows that two things have changed between 1990 and the 2017–2019 period. First, the relative size of the groups has changed over time. Currently, Hispanics are a larger share of the population than they were in 1990. Second, the income levels across groups have changed. The main counterfactual in the paper using their 1990 data equates the income of other groups to those of white individuals in 1990. The counterfactual is shown here in table 4. According to the census

Table 3. 1990 Raw Data

Group	*Share of sample*	*Average ACS labor income (incl. zeros)*	*Group-specific GDP (in trillions)*
Non-Hispanic white	0.77	$38,548	$3.8
Non-Hispanic Black	0.11	$26,617	$0.4
Hispanic	0.08	$25,144	$0.3
Other	0.04	$33,925	$0.2
Total ACS labor income			$4.6

Sources: IPUMS USA and the 1990 census.

Note: Raw data are from the 1990 census; the sample includes all individuals age 25–64 (inclusive). All dollar values are in current dollars. The last column measures the total labor income in each group as measured by the census during this period and is computed as the multiple of column 2, column 3, and 128 million individuals, the approximate BLS estimate of the aggregate population size of individuals age 25–64 during this time period.

Table 4. 1990 Counterfactual 1

Group	*Share of sample*	*Average ACS labor income (incl. zeros)*	*Group-specific GDP (in trillions)*
Non-Hispanic white	0.77	$38,548	$3.8
Non-Hispanic Black	0.11	$38,548	$0.5
Hispanic	0.08	$38,548	$0.4
Other	0.04	$38,548	$0.2
Total ACS labor income			$4.9

Source: Author's calculations.

Note: Table shows the simple counterfactual performed in the paper by equating the average labor incomes of Black and Hispanic individuals to that of white individuals.

data, equating income across groups would have increased aggregate labor income by about 7 percent (from $4.6 trillion to $4.9 trillion).

The paper concludes that the gains from equating income across groups have been rising over time. The reason, they argue, is that the 12 percent gains in 2017–2019 are larger than the 7 percent gains in 1990. However, it is worth stressing that the reason that their estimated gains are rising over time is only because population shares are shifting over time. In table 5, I perform one additional counterfactual not in the paper where I set the population weights in table 3 to those observed in 2017–2019 (as seen in table 1). This lowers imputed 1990 aggregate income to $4.4 trillion. If population shares were the same as they were in 2017–2019, equating incomes across groups in 1990 increases aggregate labor income by 12 percent (from $4.4 to $4.9 trillion); the 12 percent increase is identical to what is found in the more recent period.

CONCLUSION AND DISCUSSION My takeaways from the paper are twofold. First, the authors claim that current aggregate labor income would increase

Table 5. 1990 Counterfactual 2

Group	*Share of sample*	*Average ACS labor income (incl. zeros)*	*Group-specific GDP (in trillions)*
Non-Hispanic white	0.62	$38,548	$3.1
Non-Hispanic Black	0.12	$26,617	$0.4
Hispanic	0.17	$25,144	$0.5
Other	0.09	$33,925	$0.4
Total ACS labor income			$4.4

Source: Author's calculations.
Note: Table shows the simple counterfactual of assuming that the 1990 population shares were the same as the population shares in 2017–2019.

by about $1 trillion per year (or by 12 percent) if we equated the average incomes of Black and Hispanic individuals to the average income levels of whites individuals. Second, the authors claim that the aggregate gains have been increasing over time. But the increasing gains are solely the result of group population shares shifting over time, as I document in table 5. The income gaps across groups have been relatively constant since 1990. However, the share of Hispanic workers has been rising relative to White workers over this period. It is not that the differences in relative income are growing over time, it is just that the share of the population groups with lower income is rising.

With respect to the first takeaway, I return to my simple illustrative example at the start of my comments. The aggregate income gains from equating incomes across groups hinges critically on whether the current incomes of white individuals are a good counterfactual for Black and Hispanic individuals if group-specific barriers were removed and on how one thinks about the costs of removing such barriers. I am not sure if the counterfactual that the authors focus on is the correct one. As a result, I am not sure how seriously we should take their numbers.

Having said that, I do think that barriers to the allocation of talent across racial and ethnic groups have aggregate consequences. I believe it is good for the economics profession to have more work done exploring how large those aggregate consequences might be.

GENERAL DISCUSSION In response to the comment by Erik Hurst, Laurence Meyer argued that as the proportion of unskilled workers in the economy declines and that of skilled workers increases, the relative price of the products produced by each group will change accordingly. Low-skilled workers are likely to see their wages rise, and vice versa,

he continued, and he noted that this shift in the skill of the labor force raises allocation issues, especially related to low-skilled production and what that will look like going forward with less demand for those workers as technology changes.

Mary Daly responded by pointing out that goods and services are provided to us by the lower end of the skill distribution in other countries and that services are already increasingly automated. She concluded that the economy is very dynamic and would be able to accommodate increased educational attainment and better jobs.

In response to Hurst's comment, Caroline Hoxby noted that while there are increases in productivity following the reallocation of talent to jobs, this does not mean that there isn't also redistribution of income following decreases in discrimination, which reveals the latent distribution of productivity—an underappreciated point made by Charles Murray in *The Bell Curve*, which he primarily authored.[1] Hoxby also highlighted that we often overestimate the extent to which educational attainment has changed. Such changes have been concentrated in nonselective universities, where ethnic minorities are overrepresented, and equating these groups with the educational attainment of white workers will almost certainly overestimate what equalization would achieve because the quality of education is different, she concluded.

Daly acknowledged the redistribution issues raised by the two comments but noted that the redistribution discussed in the current paper on equalizing gaps looks at a growing pie, not a constant pie, because equalizing gaps would lead to an increase in aggregate output. She explained that while there would be a relative rewards shift, an absolute rewards shift is less likely. She argued that even if we uncover the latent distribution, it wouldn't result in nominal income cuts if the pie is increasing, but rather it is much more likely we would get a relative rebalancing. In response to the comment by Hurst, Daly agrees that change would be hard and likely be constrained by resistance to changes in the current pecking order.

Steven Davis asked why the average outcome of white individuals was the benchmark used in the paper and suggested this might need further explanation. He pointed out that there seems to be a presumption by the authors that average differences across groups are a result only of current or past discrimination when, in reality, there are other explanations as well.

1. R. J. Herrnstein and C. A. Murray, *The Bell Curve: Intelligence and Class Structure in American Life* (New York: Free Press, 1994).

15985-02b_BuckmanCom&GD-4thPgs.indd 136 5/12/22 12:09 PM

He brought up research by Raquel Fernández and Alessandra Fogli which argues that cultural differences are transmitted intergenerationally and affect labor supply and human capital investment decisions, as well as fertility preferences.[2] Davis noted that the quality-quantity trade-off in the latter has implications for human capital investment in children. Davis contends that whether such cultural differences are also a result of discrimination and barriers is an open question.

Austan Goolsbee remarked that the paper starts from the identifying assumption that if talent is equally distributed, then everybody should be able to match the distribution of white workers, which in turn would expand national income. He agrees with Hurst and Davis that taking the resulting number at face value comes with caveats and asks whether the benchmark used is the relevant one. He further notes that income increases can occur through a number of different channels, including changes in education or industry, or through the skill mix of workers—all of which would also result in a rebalancing of the general equilibrium, and, he argues, as a result some people's incomes would go down. Goolsbee then ponders whether the benchmark used in the paper is the relevant one and probes the authors for their thoughts on this.

Daly responded that the choice to use the white population was largely because it is the majority part of the population and would likely represent nondiscriminatory wages. She concedes that they could have spent more time explaining this in the paper. She pointed out that she's concerned about the cultural aspect raised by Davis and argued that it is too easy to simply accept the world today as a result of people's active choice, as opposed to acknowledging historical barriers that have been passed on intergenerationally and that ultimately become costly. Daly explained that while there may be a redistribution loss for the majority group, the point of the paper is to highlight that as the share of the population who are Hispanic increases and the share who are Black stay the same, if we don't resolve some of the current inequities, we are going to have a shrinking pie or one that grows slower as a result of fewer taxpayers and more people needing public benefits. She concluded that the point of the paper is not to leave the reader with a number but to recognize that we are essentially operating under constrained optimization where we have sidelined a lot of people in an economy which

2. Raquel Fernández and Alessandra Fogli, "Culture: An Empirical Investigation of Beliefs, Work, and Fertility," *American Economic Journal: Macroeconomics* 1, no. 1 (2009): 146–77; Raquel Fernández, "Does Culture Matter?" in *Handbook of Social Economics*, eds. Jess Benhabib, Alberto Bisin, and Matthew O. Jackson (Amsterdam: North-Holland, 2011).

15985-02b_BuckmanCom&GD-4thPgs.indd 137 5/12/22 12:09 PM

has a lower labor force participation rate than most industrialized nations.[3] Daly mentioned what she called an instrumental paper in this respect by Hsieh and others on talent allocation.[4]

Davis agreed that the reason average outcomes differ by ethnic and racial groups and between men and women is important. He said that his concern is the presumption that between-group differences are entirely due to discrimination and barriers. Without understanding what underlies those differences in their full complexity, we are unlikely to develop policies that respond effectively to them.

In the virtual conferencing chat, Hoxby noted that while it is true that the pie can increase, it is important to estimate rather than assume that all groups discriminated against could attain the benchmark.

A group of participants discussed the potential interpretation of the exercise carried out in the paper as an upper bound, with Henry Aaron arguing that the paper has as a hypothesis that white men have not drawn benefits from historical discrimination against women and Black people, which he finds absurd. Marianne Bertrand added that it may not be an upper bound if diversity in the labor force and in workplaces fosters more innovative ideas, which could effectively extend the frontier. Michael Kiley noted that, related to the potential productivity effects that greater equality could bring about, there is research suggesting that higher incomes are associated with a greater propensity to innovate.[5] Aaron suggested that while there is no doubt that removing barriers to the expression of talent would bring enormous economic, social, and political gains, the measure in the paper, which uses the earnings of white men with the legacies of discrimination that have led to the current conditions, is not a good guide to the actual magnitude of potential gains.

Kiley mentioned future infrastructure investments and suggested a role for the reallocation of resources, including jobs, rather than more resources. He pointed to the importance of the macroeconomic backdrop for structural reallocation.

3. Mary C. Daly, Joseph H. Pedtke, Nicholas Petrosky-Nadeau, and Annemarie Schweinert, "Why Aren't U.S. Workers Working?," *FRBSF Economic Letter*, November 13, 2018, https://www.frbsf.org/economic-research/publications/economic-letter/2018/november/why-are-us-workers-not-participating/.

4. Chang-Tai Hsieh, Erik Hurst, Charles I. Jones, and Peter J. Klenow, "The Allocation of Talent and U.S. Economic Growth," *Econometrica* 87, no. 5 (2019): 1439–74.

5. Alex Bell, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," *Quarterly Journal of Economics* 134, no. 2 (2019): 647–713.

15985-02b_BuckmanCom&GD-4thPgs.indd 138 5/12/22 12:09 PM

Harry Holzer asked how much the gap would increase if one could take into account the underrepresentation of Black men in the CPS data, mostly due to incarceration and other undercounting issues.

Ricardo Reis brought up a paper by Michael Clemens that describes a similar exercise to the current paper and finds a gain equal to 50–150 percent of GDP resulting from the average immigrant being able to move without barriers to rich countries like the United States—which would benefit world equity enormously.[6]

Robert Gordon commented that if we raise the average income for Black people by having more Black CEOs, there would be fewer white CEOs, reducing the benchmark average white income. He suggested that the difference between Black male and female employment rates is partly due to Black male incarceration and that the paper needs to indicate whether all Black prisoners are to be released and how to deal with the employment stigma of those with felony convictions. Stating that only about one-quarter of Black students in central city high school districts pass reading and math tests at the eleventh-grade level, Gordon concluded that this has many consequences: high school dropouts, more Black students in community colleges than in four-year colleges, and more in low-quality, nonselective four-year colleges, ultimately making it unrealistic to expect Black workers to achieve equal earnings with white workers.

Luigi Zingales pointed out that if we look back at projections economists did in the past, starting with Malthus, about the future path of the economy, we can get a good sense of the true uncertainty about our forecasts—a humbling exercise.

6. Michael A. Clemens, "Economics and Emigration: Trillion-Dollar Bills on the Sidewalk?," *Journal of Economic Perspectives* 25, no. 3 (2011): 83–106.

15985-02b_BuckmanCom&GD-4thPgs.indd 140 5/12/22 12:09 PM

SUSAN CHERRY
Stanford University

TOMASZ PISKORSKI
Columbia University

ERICA JIANG
University of Southern California

AMIT SERU
Stanford University

GREGOR MATVOS
Northwestern University

Government and Private Household Debt Relief during COVID-19

ABSTRACT We study the suspension of household debt payments (debt forbearance) during the COVID-19 pandemic. Between March 2020 and May 2021, more than 70 million consumers with loans worth $2.3 trillion entered forbearance, missing $86 billion of their payments. This debt relief can help explain the absence of consumer defaults relative to the evolution of economic fundamentals. Borrowers' self-selection is a powerful force in determining forbearance rates: relief flows to households suffering pandemic-induced shocks that would otherwise have faced debt distress. Moreover, 55 percent of forbearance is provided to less creditworthy borrowers with above median income and higher debt balances—that is, those excluded from income-based policies, such as the stimulus check program. A fifth of borrowers in forbearance continued making full payments, suggesting that forbearance acts as a credit line. By May 2021, about 60 percent of borrowers had already exited forbearance while more financially vulnerable and lower income borrowers were still in forbearance with an accumulated debt overhang of about $60 billion. Exploiting a discontinuity in mortgage eligibility under the CARES Act, we estimate that implicit government debt relief subsidies increase the rate of forbearance by about a third. Government relief is provided through private intermediaries, with shadow banks less likely to provide forbearance than traditional banks.

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

15985-03a-Cherry-4thPgs.indd 141 5/12/22 12:09 PM

Large economic crises such as the Great Depression and Great Recession are often accompanied by significant household debt distress, which spills over to the rest of the economy (Mian and Sufi 2009; Keys and others 2013; Wachter, Duca, and Popoyan 2019). Based on historical experience, the evolution of economic fundamentals of the COVID-19 pandemic would also predict a large amount of household debt distress. This grim scenario failed to materialize, resulting in substantial "missing defaults." We study the role of private and government debt forbearance—that is, temporary suspension of debt repayments—in averting household debt distress. We find that forbearance can explain a significant part of the missing household defaults and likely significantly dampened the potential negative spillovers to the rest of the economy. Borrower self-selection and take-up played a central role in the incidence and effectiveness of the relief. Government mandates and intermediary factors also played an important role in transmission of relief. We also discuss how to unwind the debt accumulated by vulnerable households and draw broader implications for the design of debt relief policies.

We study forbearance using a representative credit bureau panel of more than 20 million US consumers. The data allow us to study which loans are in forbearance—allowing borrowers to defer loan payments—as well as the extent to which households chose to miss payments. A significant share of households, for example, request and obtain forbearance but nevertheless continue making full payments. The data also allow us to classify which loans were eligible for government debt relief under the Coronavirus Aid, Relief, and Economic Security (CARES) Act and which relief was provided by the private sector.

Over 70 million individuals obtained forbearance between March 2020 and May 2021, totaling loans worth about $2.3 trillion. The lion's share of new forbearance initiated during the COVID-19 crisis was in the categories of mortgages and student debt, accounting for $1.4 trillion and $655 billion, respectively. Forbearance actions resulted in substantial financial relief for households. The average cumulative payments missed by individuals in forbearance during this period were largest for mortgage ($4,254) and auto ($398) debt. By May 2021, debt forbearance allowed US consumers to miss about $86 billion of their payments. At this rate, forbearance would allow more than 70 million consumers to miss about $100 billion of their debt payments by the end of September 2021, when some of the key government forbearance mandates were set to expire.

The extent of forbearance may account for the missing household defaults during the pandemic. Economic fundamentals deteriorated significantly

15985-03a-Cherry-4thPgs.indd 142 5/12/22 12:09 PM

during the pandemic, with the unemployment rate reaching almost 15 percent in 2020:Q2. The strong historical association between unemployment and mortgage default predicts a substantial increase in household debt distress (Piskorski and Seru 2018). Instead, delinquency rates declined from 3 percent to 1.8 percent. Exploiting the richness of our individual-level panel data, we measure missing defaults. We estimate the expected delinquency levels, given the evolution of the local economic conditions and the credit profile of borrowers, and compare them to actual levels that occurred during the pandemic. Our estimates suggest that the majority of predicted mortgage defaults are missing. The actual default rate averaged below 2 percent instead of a predicted 6.8 percent at its peak, amounting to about 1.5 to 2.5 million missing defaults in the aggregate. While other policies such as generous unemployment benefits certainly played a role in averting consumer distress (Cox and others 2020), a back-of-the-envelope calculation suggests that the level of forbearance is large enough to account for averted potential delinquencies in the mortgage market. Moreover, despite its much lower cost, we find that the extent of forbearance relief is much more strongly related to the extent of missing defaults in a region than other stimulus programs. We further validate this view by exploiting government mandates that generate variation in the forbearance rates among similar borrowers and show a strong association between forbearance and missing default rates. We speculate that the resultant low delinquencies can explain, at least in part, why the pandemic has not resulted in house price declines, which would have further exacerbated household debt distress.

There are at least two features that distinguish household debt forbearance from other relief programs targeted at households. First, borrowers self-select into forbearance, as well as decide whether to draw on the forbearance, which is effectively a line of credit. We show that this self-selection is an important determinant of how debt relief is allocated in the population, and forbearance provides a temporary bridge for pandemic-related liquidity shocks faced by the households. Second, the private sector plays an important role in the provision of forbearance, both as an alternative to government forbearance and as a conduit through which government forbearance is implemented.

To obtain forbearance, borrowers must request it from the lender, and in the case of private forbearance, lenders must approve such requests. Among the largest consumer debt category, residential mortgages, more than 90 percent of borrowers eligible for forbearance through the CARES Act decided not to take up the option of debt relief. This suggests that borrowers' self-selection is a powerful force in determining forbearance

rates. This self-selection resulted in relief being provided to a very different population of individuals relative to other CARES Act policies, such as stimulus checks. The rates of forbearance also decline substantially with creditworthiness but are much less progressive in income than other relief programs. Lower income households are more likely to obtain forbearance relief. Because they have lower debt balances, the dollar value of debt relief is also smaller. About 55 percent of the dollar amount of financial relief from forbearance was received by borrowers with above median prepandemic incomes. This observation highlights an important feature of forbearance: it provides relief to borrowers with higher prepandemic incomes who may become financially constrained during the pandemic but who do not qualify for income-based relief programs. Notably, such individuals can play an important role in aggregate responses due to their high marginal propensities to consume.

We provide further evidence that self-selection into forbearance also provides relief to households suffering pandemic-induced shocks who would have otherwise faced debt distress—the population that is potentially targeted by the policy. Forbearance rates are significantly higher in regions that experienced the highest COVID-19 infection rates and the greatest deterioration in their local economies, as reflected by unemployment insurance claims and the concentration of industries most exposed to the pandemic. The economic and health consequences of the pandemic have disproportionately affected minorities, especially Black Americans (Lopez, Rainie, and Budiman 2020). Consistent with this observation, regions with higher shares of minorities received debt forbearance at higher rates. We also document the highest rates of forbearance in regions where economic conditions would otherwise predict the highest default rates on household debt. These are regions where we also observe the largest gap between expected and actual defaults. Thus forbearance may have reached its intended target, especially helping households who were likely affected by the pandemic but were unlikely to be eligible for income-based programs.

The private sector provided forbearance for debt outside the federally insured mortgages and student loans mandated by the CARES Act. We find substantial increases in forbearance in auto and credit card loans, as well as mortgage loans not eligible under the CARES Act: about 20 percent of total debt relief was provided by the private sector for debt not eligible under CARES Act rules.

We compare the provision of private and public sector forbearance to measure the role of implicit forbearance subsidies provided in the government mandate. Private forbearance is presumably the result of (ex ante) a

mutually beneficial renegotiation, which allows borrowers to bridge a temporary liquidity shock. To evaluate the importance of implicit government subsidies, we exploit a size discontinuity in the eligibility of mortgages for relief under the CARES Act. While government-insured loans below the conforming loan limit qualified for government-mandated forbearance, loans above the limit were not eligible.[1] Restricting our analysis to mortgages with balances near the conforming loan balance limits, we find that the percentage of loans in forbearance increases by 1.6 percent, about a third in relative terms, for loans covered by the government mandate. This contrasts with the pattern observed outside of the pandemic, in which loans issued without government guarantees have slightly higher forbearance rates at the discontinuity. Our back-of-the-envelope estimates suggest that about 25 percent of government forbearance is subsidized and the rest is provided to borrowers who might have received debt relief from the private sector.

The estimates are likely a lower bound for various reasons. First, the government mandates might have affected private forbearance supply. This positive spillover could be generated through several channels. For example, the CARES Act sets uniform forbearance protocols and spurred a collective action response that might not have occurred so promptly otherwise. The standards set by the CARES Act might have not only provided servicers with simple rules for the private sector to follow but also imposed reputational concerns on servicers who did not supply forbearance for loans not covered by the mandates. Second, the government mandates might have affected loans not covered by the mandates through general equilibrium: the mandates avoided delinquencies and costly foreclosures and could stabilize house prices, which prevented loans collateralized by the houses and not covered by the mandates from going underwater (Anenberg and Scharlemann 2021).

We find evidence that this additional forbearance seems to decrease household distress relative to predicted levels based on economic fundamentals. A 1.6 percent higher forbearance rate during the pandemic on loans covered by mandates is associated with a 0.7 percent higher rate of missing defaults. These estimates imply that two forbearances are associated with about one missing default, the same ratio as in the aggregate data. This further validates our observation that debt forbearance can account for a substantial portion of prevented defaults during the pandemic.

1. Jumbo loans exceed the conforming loan balance limits set by the Federal Housing Finance Agency and cannot be purchased, guaranteed, or securitized by the government-sponsored enterprises (GSEs).

Government relief is explicitly provided by a variety of private servicers, more than half of whom are shadow banks (Buchak and others 2018; Jiang 2020; Jiang and others 2020). Since relief of government loans is mandated, one might expect that there would be few differences between suppliers. Instead, even accounting for borrower characteristics, we find lower rates of forbearance for loans serviced by shadow banks relative to traditional banks. This result suggests that despite the blanket (and relatively simple) government mandate, who implements forbearance has a meaningful effect on the amount of debt relief provided—a result reminiscent of debt relief during the Great Recession (Agarwal and others 2017; Agarwal and others 2020).

One of the problems faced by policymakers is that it is difficult to recognize which households need to bridge temporary liquidity shocks and which households suffer more permanent debt distress, leaving them insolvent. This is complicated by the fact that the policy intervention itself can affect the duration of the crisis. Forbearance is designed as a temporary bridge to absorb liquidity shocks faced by households—deferred payments need to be repaid. In fact, we document that more than 20 percent of households obtain forbearance but nevertheless continue making full payments. These are households who behaved as if forbearance were a line of credit that they could draw on in need, but who realized ex post that they did not need to access it. On the other end of the spectrum are borrowers who are insolvent and who will not be able to exit forbearance without significant loan modification. An important policy question is therefore how forbearance will be unwound after it expires; government mandates issued during the pandemic were to expire on September 30, 2021.

We first document that a substantial share of borrowers who entered forbearance did so to bridge temporary shocks, but we also document a substantial amount of forbearance overhang of postponed payments for a significant share of borrowers. About 60 percent of borrowers had already exited forbearance as of May 2021 (about 75 percent of mortgage borrowers). Most of these borrowers used forbearance as a temporary liquidity facility—either not drawing down on payments (a third) or repaying missed payments within two months of entering forbearance (about 20 percent). On the other hand, a significant proportion of borrowers (about 7 percent) who exited forbearance did so with a loan modification, suggesting that their distress was not temporary.

In addition, at the time of writing a substantial share of borrowers had not yet exited forbearance: as of May 2021 more than 40 percent of 72 million Americans who entered forbearance during the pandemic were still missing

15985-03a-Cherry-4thPgs.indd 146 5/12/22 12:09 PM

about $60 billion on their debt repayments.[2] At this rate, by September 2021 when forbearance mandates are set to expire, we estimate that these borrowers with persistent periods of forbearance will be left with a forbearance overhang of more than $70 billion in accumulated postponed repayments. This estimated overhang amounts to about $3,900 per individual, which is about 1.5 times their average monthly income and more than 2.2 times for lower income borrowers. For mortgage borrowers, the largest debt category, the estimated overhang is about $15 billion, amounting to about $14,200 per individual on average, which is about 3.4 times their average monthly income. Moreover, as discussed above, these borrowers with long periods of forbearance are more likely to be in regions with lower income, higher unemployment, and higher minority share.

A significant share of these low-income borrowers will likely become distressed if accumulated payments are structured as a onetime payment due immediately after forbearance ends, even if this payment is anticipated. Most mortgages in forbearance, including the ones held by the most vulnerable borrowers, are insured through the government-backed programs, allowing wide latitude in implementation. Adding missed payments to the loan balance would spread out the repayment of payments in forbearance over a long period of time (about 25 years), increasing existing payments by about $90–120 per month. In addition, the government could consider a refinancing program that would allow borrowers in forbearance to easily refinance their loans while increasing the loan balance of the new loan by the accumulated amount of missed payments in forbearance. Such a program could be part of a broader refinancing initiative (Golding and others 2020). Since borrowers in forbearance face mortgage rates considerably higher than the current rates (in excess of 4 percent on legacy loans), refinancing could lower the overall mortgage payment burden of borrowers. The upfront versus deferred repayments could have significantly different consequences for consumers and for the aggregate economy (Eberly and Krishnamurthy 2014; Mian and Sufi 2014b; Piskorski and Seru 2018; Ganong and Noel 2020).

We conclude by drawing broader implications for debt relief policies. One possible reason for the quick implementation of debt relief actions during COVID-19 is that the private sector and policymakers may have internalized the lessons from the Great Recession pointing to significant

2. Borrowers who entered forbearance March–October 2020 missed about $43.5 billion of their debt payments. Accounting for debt repayments already made by October 2020, the net amount is $38 billion.

costs of widespread defaults and foreclosures and were more willing to provide widespread and quick debt relief (Eberly and Krishnamurthy 2014; Campbell, Clara, and Cocco 2020; Piskorski and Seru 2018). The large private response suggests that a substantial amount of debt forbearance was mutually beneficial. Another alternative reason for such behavior could be that the COVID-19 shock was perceived as more transitory relative to prior crises, which could have promoted a more widespread deployment of temporary debt relief measures by the private sector. This is consistent with the consumer debt design literature, which indicates that lenders should provide a certain amount of debt relief during economic downturns to limit deadweight costs of default and allow better risk-sharing between borrowers and lenders, especially if the underlying shocks are transitory (Piskorski and Tchistyi 2010, 2011, 2017; Eberly and Krishnamurthy 2014; Greenwald, Landvoigt, and Van Nieuwerburgh 2017; Guren, Krishnamurthy, and McQuade 2021; Campbell, Clara, and Cocco 2020; Ganong and Noel 2020). Relatedly, the COVID-19 shock is a textbook example of a rare aggregate exogenous shock that is largely outside of agents' influence. This should alleviate concerns about the moral hazard effects of debt relief on incentives to repay debt leading to more widespread loan renegotiation efforts during such times (Piskorski and Tchistyi 2010, 2011, 2017; Mayer and others 2014).

We note that our results also suggest that allowing borrowers a choice of whether to request debt relief, as in the case of mortgages, might have resulted in a potentially better targeted debt relief policy as compared to blanket automatic forbearance policies like the one used in the case of student debt. However, policies employing such self-selection can still expose borrowers to intermediary related implementation frictions. As we show these frictions were still present during the pandemic despite the significantly simpler design of debt relief policies relative to those used in the Great Recession. This suggests that future debt relief policies leaning on borrower self-selection for better targeting may also need to account for possible intermediary frictions. Finally, because most of the forbearance amounts will likely be repaid, unlike other stimulus measures forbearance implies substantially smaller net transfers to agents. Yet despite their much lower cost to taxpayers, the targeting of temporary relief at households in distress (through self-selection) prevented substantial household distress and with it likely the spillover to the rest of the economy.

Our paper is related to the literature on the role of the household balance sheet channel in the transmission of economic shocks (Mian and Sufi 2009,

2011, 2014a; Guerrieri and Uhlig 2016; Hurst and others 2016; Agarwal and others 2017, 2018, 2020; Berger and others 2018, 2021; Benmelech, Meisenzahl, and Ramcharan 2017; Kaplan, Mitman, and Violante 2020; Favilukis, Ludvigson, and Van Nieuwerburgh 2017; Di Maggio and others 2017; Di Maggio, Kermani, and Palmer 2020; Greenwald 2016; Guren and others 2018; Auclert 2019; Beraja and others 2019; Eichenbaum, Rebelo, and Wong 2019; Andersen and others 2020). Within this literature our paper contributes to the recent studies that analyze the effects of various stabilization programs operating through the household balance sheet channel (Mian and Sufi 2012; Parker and others 2013; Hsu, Matsa, and Melzer 2018; Berger, Turner, and Zwick 2020) and especially studies focusing on various forms of debt relief (Piskorski, Seru, and Vig 2010; Agarwal and others 2011, 2017, 2020; Mayer and others 2014; Scharfstein and Sunderam 2016; Di Maggio and others 2017; Di Maggio, Kermani, and Palmer 2020; Maturana 2017; Fuster and Willen 2017; Kruger 2018; Piskorski and Seru 2018, 2021; Auclert, Dobbie, and Goldsmith-Pinkham 2019; Mueller and Yannelis 2019b; Ganong and Noel 2020). It is also related to literature on effects and policy response to the pandemic (Baker and others 2020; Chetty and others 2020; Coibon, Gorodnichenko, and Weber 2020; Cox and others 2020; Elenev, Landvoigt, and Van Nieuwerburgh 2020; Granja and others 2020; Guerrieri and others 2020; Fuster and others 2021).

I. Institutional Setting: US Consumer Debt Market, Debt Forbearance, and the CARES Act

A forbearance agreement includes a halt or reduction in a borrower's loan payments for a fixed period. To enter a forbearance agreement the borrower must usually approach the lender with satisfactory proof of distress and proof that the distress is temporary. If the lender chooses to extend forbearance, the borrower may stop or decrease their loan payments without fear of foreclosure, or they can keep making contractual payments. Although the payments have been delayed, the loan's interest does not stop accruing over this period. Forbearance is not a debt forgiveness program or a loan modification. The borrower is required to pay the lender the missed payments after the forbearance period ends. Typical repayment plans following the end of the forbearance period consist of a lump sum payment or increasing the regular payment amounts once forbearance is finished. Borrowers can also attempt to obtain a loan modification agreement that allows them further deferral or permanent reduction of at least part of their missed payments.

The CARES Act signed on March 27, 2020, included several loan forbearance provisions.[3] Below, we discuss typical structures and standards for loan forbearance in each of the loan segments and the implications of the CARES Act for them. We focus on the four main categories of consumer debt: residential mortgages, auto, revolving, and student debt. Figure A1 in the online appendix shows the evolution of the aggregate outstanding loan balance for these four types of debt from 2006 to 2021.

I.A. Residential Mortgage Market

The residential mortgage market is by far the largest form of consumer debt in the United States. As of 2020, total mortgage balances in the United States totaled about $10 trillion. Roughly two-thirds of outstanding mortgages are effectively guaranteed by the US government (Buchak and others 2018) comprising conforming loans sold to government-sponsored enterprises (GSEs) and the Federal Housing Administration (FHA) loans. Conforming loans are typically extended to borrowers with relatively high credit scores, fully documented income and assets, and moderate loan-to-value ratios. The FHA provides mortgage insurance on loans made by FHA-approved lenders nationwide, which is usually considered the riskiest segment of the mortgage market as they mainly appeal to lower income and less creditworthy households. GSE and FHA loans are subject to origination loan balance limits.[4] Jumbo loans with balances exceeding the conforming loan limit account for about 15 percent of the outstanding loan balances in our data. These loans are ineligible for government guarantees and are therefore much more difficult to securitize and are typically retained on lender balance sheets (Buchak and others 2020).

Mortgage forbearance agreements have been historically granted on a case-by-case basis (Corcoran and Haltom 2020). Proof of distress, proof that the distress is temporary, and proof that the borrower can repay the interest and missed payments has been typically required when granting forbearance. Under the CARES Act, borrowers with federally backed mortgages, including GSE, FHA, and Veterans Administration loans, were allowed to pause their mortgage payments with no penalties until

3. US Department of the Treasury, "Covid-19 Economic Relief," https://home.treasury.gov/policy-issues/coronavirus.

4. For conforming loans, these limits were $417,000 for a single-family home in most regions and $625,000 in some regions depending on the area's cost of living throughout most of the last decade, reaching $510,400 and $765,600, respectively, by 2020 after their progressive increases during the 2017–2020 period.

September 2021. Under forbearance policies, no fees, penalties, or additional interest may be added to a borrower's account. In addition, the CARES Act granted mortgage borrowers protections to help them avoid foreclosure, including a sixty-day foreclosure and eviction moratorium for borrowers with federally backed mortgages. This moratorium was originally intended to expire in May 2020 but was subsequently extended several times.[5] Borrowers with loans without the government guarantees, such as jumbo loans, were not covered by the CARES Act forbearance mandates, in which case forbearance had to be approved by the lender.[6]

It is important to note that while the CARES Act guaranteed individuals with federally backed mortgages the right to pause their mortgage payments, it did not automatically place their mortgages in forbearance. Borrowers were required to contact their loan servicer to put their payments on hold, though the forbearance process was straightforward—borrowers simply needed to claim they had a pandemic-related hardship and did not need to submit any documentation.

I.B. Student Debt

In 2006, student loan debt was the smallest of the four consumer debt categories, but by 2020 total student loan balances were exceeded only by mortgages (online appendix figure A1). The federal government is the primary provider of student loans in the United States, with about 90 percent of outstanding student loans held by the Department of Education (Looney and Yannelis 2019). Obtaining federal student loan forbearance has historically been relatively easy. For example, federal student loan borrowers may be entitled to a loan deferment if they are unemployed or a forbearance if the amount owed exceeds 20 percent of their gross income (Mueller and Yannelis 2019a).

The CARES Act automatically placed federal student loans into administrative forbearance and set their interest rate to 0 percent. Student loan forbearance was originally designed to expire in September 2020 but has been extended several times. Borrowers with private student loans are not covered under the CARES Act, so forbearance had to be approved by the lender in those cases.

5. For a discussion of the mortgage foreclosure moratorium, see Capponi, Jia, and Rios (2020).

6. The act prevents the reporting of delinquency to credit bureaus on all loans if the borrower is current on their account and their lender agrees to allow a skipped payment, partial payment, or other accommodation due to the pandemic.

I.C. Auto and Revolving Debt

Auto debt has increased faster than all other types of debt except for student loans over the past ten years, totaling over $1.3 trillion in 2020. The majority of both new and used cars are financed with debt—in 2018, over 85 percent of new cars and 55 percent of used cars were financed with a loan or a lease (Di Maggio and others 2017). Revolving debt refers to all accounts that allow individuals to borrow against a credit line, except for home equity lines of credit. This includes credit cards as well as retail and other revolving accounts like personal lines of credit (Agarwal and others 2018). Total outstanding revolving debt was slightly over $1 trillion as of 2020 (online appendix figure A1). Unlike the mortgage and student loan markets, the CARES Act did not include any explicit forbearance mandates for auto or revolving loans. Borrowers contacted their lenders for information about forbearance or deferment policies for these types of debt, with policies and eligibility varying by lender.

II. Data Sources

II.A. Equifax Analytic Dataset

Our main data set is the Analytic Dataset provided by Equifax. Equifax is a credit-reporting agency that provides monthly borrower-level data on credit risk scores, consumer age, geography, debt balances, and delinquency status at the loan level for all consumer loan obligations and asset classes. The Analytic Dataset is created from a 10 percent random sample of the US credit population from 2005 to 2021 across the United States and consists of over 20 million consumers (table 1). Randomization in the sample is based on Social Security numbers, ensuring that the sample is representative of the US credit population. We use these data to investigate consumer forbearance status, delinquency status, payment history, age, income, credit score, and location.

We follow Equifax's standard procedure for identifying whether a loan is in forbearance. Specifically, we consider a loan to be in forbearance if it has a narrative code indicating that it is in forbearance or deferment, if it is in a partial payment plan, if the loan has been modified, or if the account has a positive balance with no reported scheduled payment. We exclude all loans that have been refinanced or prepaid. To validate the forbearance measure, we compare a subset of the Equifax sample labeled as GSE loans to loans in the Fannie Mae Single-Family Historical Loan Performance Dataset, which we describe in detail below. The samples are comparable in

15985-03a-Cherry-4thPgs.indd 153

Table 1. Individual-Level Descriptive Statistics

	All individuals				*All individuals in forbearance*			
	Pre-COVID-19		*COVID-19*		*Pre-COVID-19*		*COVID-19*	
	(1) *Mean*	*(2)* *SD*	*(3)* *Mean*	*(4)* *SD*	*(5)* *Mean*	*(6)* *SD*	*(7)* *Mean*	*(8)* *SD*
Vantage	701	122	711	119	645	122	670	112
Estimated age	49	18	49	18	41	17	40	16
Estimated income	39,992	19,272	40,134	19,426	34,063	19,250	33,940	18,562
Fraction with first mortgage	0.22		0.22		0.18		0.21	
Fraction with auto debt	0.35		0.35		0.38		0.40	
Fraction with revolving debt	0.89		0.93		0.78		0.74	
Fraction with student debt	0.18		0.17		0.56		0.64	
Average first mortgage balance (nonzero accounts)	210,020	240,631	216,969	248,378	232,542	284,857	236,257	266,933
Average combined mortgage balance (nonzero accounts)	204,346	239,936	218,478	247,674	224,272	283,478	229,115	266,321
Average auto debt (nonzero accounts)	18,474	17,496	19,605	17,898	19,897	20,567	20,173	19,715
Average revolving debt (nonzero accounts)	5,047	9,178	6,077	10,109	7,369	12,370	6,482	11,389
Average student debt (nonzero accounts)	35,149	51,510	36,198	52,449	39,234	54,999	37,716	53,975
Average credit card utilization (%)	32		27		49		42	
Average available credit	17,512	23,338	17,776	23,488	14,993	25,582	15,252	24,692
Number of consumers	20,265,012		21,200,723		3,702,943		9,338,267	

Source: Equifax.

Note: This table shows summary statistics for all individuals in the Equifax data set (columns 1 through 4) and all individuals in forbearance (columns 5 through 8). Pre-COVID-19 statistics are based on January and February 2020, while COVID-19 statistics are based on March 2020 to May 2021. The data from Equifax comprise a 10 percent representative sample of the US credit population.

5/12/22 12:09 PM

terms of borrower and loan characteristics (online appendix table A1) as well as forbearance rates (online appendix figure A2). If anything, we underestimate the amount of forbearance using Equifax data.

II.B. Fannie Mae Single-Family Historical Loan Performance Dataset

We obtain all loans that were acquired by Fannie Mae since January 1, 2000, and the monthly performance through October 2020. We restrict our sample to active loans, which had not been paid off, refinanced, or foreclosed by January 2020. These loan-level monthly panel data provide detailed information on a rich set of loan and borrower characteristics (e.g., FICO scores, loan-to-value, debt-to-income, location of the property, and interest rates), property, and monthly payment history. Two sets of information are important to supply our forbearance analysis.

First, we can determine if a seller or servicer is a bank or a shadow bank by merging the Fannie Mae data set to bank regulatory filings (e.g., Forms 031 and FY-9C) and shadow bank call reports (Jiang and others 2020) for entities that represent at least 1 percent of volume within a given acquisition or reporting quarter. Second, Fannie Mae collects information about the type of assistance plan that the borrower is enrolled in that "provides temporary mortgage payment relief or an opportunity for the borrower to cure a mortgage delinquency over a defined period," in which Forbearance Plan, Repayment Plan, and Trial Period Plan are the three major borrower assistance plan categories.[7]

II.C. Other Data Sources

The Opportunity Insights Tracker provides real-time data on total COVID-19 case rates, total unemployment insurance claims, changes in credit/debit card spending, and changes in time spent at workplaces at the county level.[8] We use the data to understand how local economic conditions and regional impacts of the COVID-19 crisis relate to forbearance actions. We supplement the regional Opportunity Insights Tracker with socioeconomic characteristics from the US Census Bureau's American Community Survey 2018 five-year estimates at the zip code tabulation area level.[9] We use median house prices from Zillow, unemployment claims and benefits

7. Fannie Mae, Single-Family Loan Performance Dataset and Credit Risk Transfer—Glossary and File Layout, https://capitalmarkets.fanniemae.com/media/6931/display, p. 8.

8. Opportunity Insights, "Economic Tracker," https://opportunityinsights.org/tracker-resources/.

9. US Census Bureau, "American Community Survey Data," https://www.census.gov/programs-surveys/acs/data.html.

from the Department of Labor, the number and amount of economic impact payments (stimulus checks) received by each state from the Internal Revenue Service, and the number and size of Paycheck Protection Plan loans from the Small Business Administration.[10] We also gather information on the number of small businesses in a county and the share of the workforce employed in certain industries from the Bureau of Labor Statistics.[11]

III. Aggregate Household Debt Forbearance and the Absence of Distress during COVID-19

III.A. Aggregate Forbearance Rates, Usage, and Amount of Relief

We begin by analyzing forbearance rates on residential mortgages, the largest category of US consumer debt. Residential mortgage forbearance rates increased from roughly 0.6 percent prior to the COVID-19 pandemic to nearly 7 percent in June following the declaration of the national COVID-19 emergency and the implementation of the CARES Act in March 2020 (figure 1, panel B). Overall, about 9.45 percent of mortgage borrowers were in forbearance during the period from March 2020 to May 2021, of which about 90 percent entered forbearance during the COVID-19 period (table 2, panel A). Forbearance rates during the COVID-19 period were also much larger than those during the Great Recession, when forbearance rates peaked at a little over 2 percent.

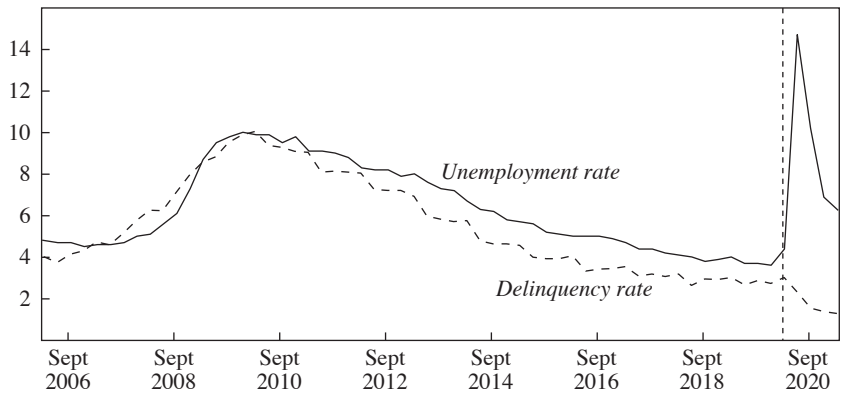
Figure 2 shows delinquency and forbearance rates of different debt types from January 2019 to May 2021. Auto and revolving debt exhibited low forbearance rates prior to the pandemic, with large spikes occurring around April 2020. Student loans are the exception, with large numbers of loans in forbearance or deferment prior to the pandemic (about 50 percent). Nevertheless, the percentage of student loans in forbearance or deferment jumps from 50 percent to well over 90 percent in April 2020. Since June 2020, forbearance rates have declined for all debt types except for student loans, which were placed in automatic forbearance, but they remain elevated well above their historical averages.

10. Zillow, "United States Home Values," https://www.zillow.com/research/data/; US Department of Labor, "Monthly Program and Financial Data," https://oui.doleta.gov/unemploy/claimssum.asp; Internal Revenue Service, "SOI Tax Stats—Coronavirus Aid, Relief, and Economic Security Act (CARES Act) Statistics," https://www.irs.gov/statistics/soi-tax-stats-coronavirus-aid-relief-and-economic-security-act-cares-act-statistics#EIP3; US Small Business Administration, "PPP Data," https://www.sba.gov/funding-programs/loans/covid-19-relief-options/paycheck-protection-program/ppp-data.

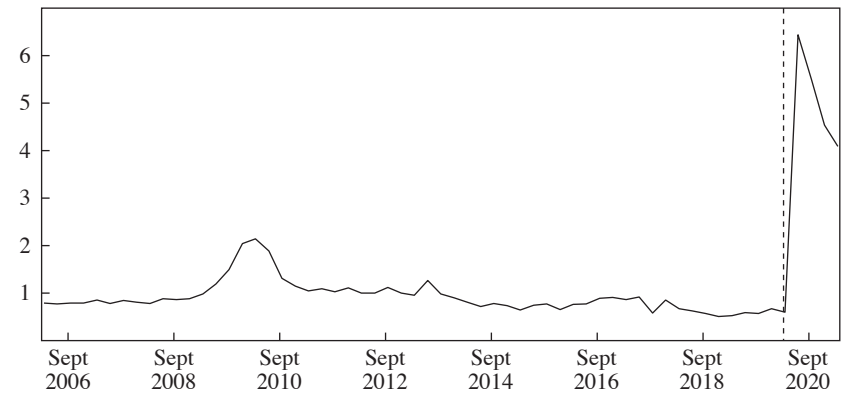
11. US Bureau of Labor Statistics, "County Business Patterns: 2019," https://www.census.gov/data/datasets/2019/econ/cbp/2019-cbp.html.

Figure 1. Residential Mortgage Debt: Delinquency and Forbearance Rates

Panel A: Delinquency and unemployment rate



Panel B: Forbearance rate and the COVID-19 crisis



Sources: Equifax; US Bureau of Labor Statistics.

Note: Panel A shows the US residential mortgage delinquency rate along with the US unemployment rate over the time period from Q2:2006 to Q1:2021. Panel B shows the corresponding mortgage forbearance rates. Delinquency rates are thirty days past due or worse. The dashed vertical line shows the declaration of national emergency due to COVID-19 and the passage of the CARES Act. Forbearance rates are calculated according to Equifax's standard procedure for identifying loans in forbearance. The quarterly unemployment rates are peak values in a respective quarter. The data from Equifax comprise a 10 percent representative sample of the US credit population.

15985-03a-Cherry-4thPgs.indd 156 5/12/22 12:09 PM

Table 2. Consumer Debt Forbearance during COVID-19

Panel A: Mean delinquency and debt forbearance rates during the COVID-19 period

	Mortgages	*Auto loans*	*Revolving loans*	*Student loans*
Monthly delinquency rate (%)	1.52	2.92	1.84	0.78
Monthly forbearance rate (%)	4.34	2.32	2.38	85.21
Missing payments/in forbearance (%)	74.50	77.05	81.39	96.22
Scheduled payment missed/ in forbearance (%)	63.99	63.91	83.07	92.47
Monthly amount missed per individual/in forbearance ($)	$1,084.7 [$3,008.70]	$253 [$324]	$32.84 [$114.30]	$28.96 [$93.09]
Ever in forbearance (%)	9.45	10.50	4.63	93.03
Entered forbearance (%)	8.07	7.70	10.19	34.22
Cumulative amount missed per individual/in forbearance ($)	$4,254 [$14,604]	$398 [$820]	$60.7 [$428.6]	$312 [$1,094]

Panel B: Aggregate debt forbearance statistics

	Mortgages	*Auto debt*	*Revolving debt*	*Student debt*
Loan balance in forbearance (%)	8.89	7.60	9.34	44.5
Amount of loans in forbearance ($ billion)	1,400	198	125	655
Number of loans in forbearance (million)	6.3	11	62	68
Missed in forbearance (average per month) ($ billion)	2.6	0.331	0.392	3.7
Missed in forbearance (March–May 2021) ($ billion)	31	5.7	4.7	45

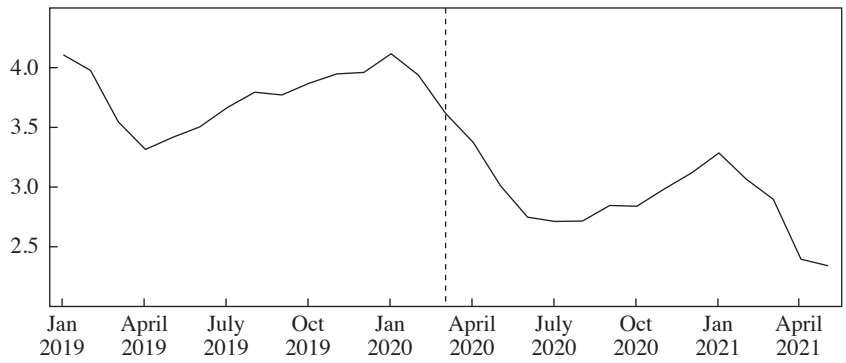
	All debt types
Number of loans that entered forbearance (million)	147
Number of individuals who entered forbearance (million)	72

Source: Equifax.

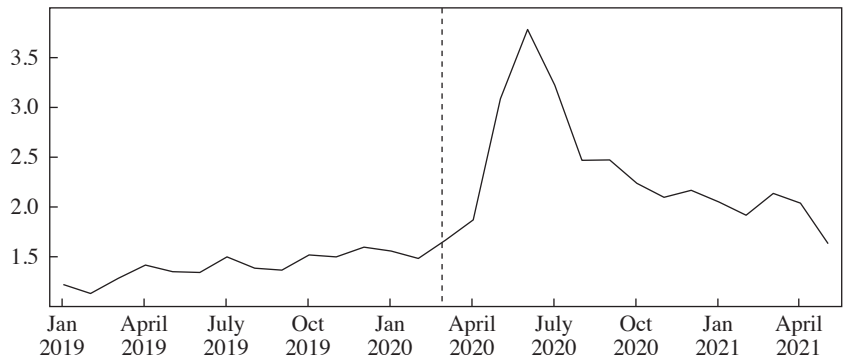
Note: Panel A shows summary statistics related to forbearance for different debt types during the COVID-19 period (March 2020 through May 2021). Standard deviations shown in brackets. Panel B provides aggregate statistics on loans that entered forbearance during the COVID-19 period. The data from Equifax comprise a 10 percent representative sample of the US credit population.

Figure 2. Auto, Revolving, and Student Debt: Delinquency and Forbearance Rates

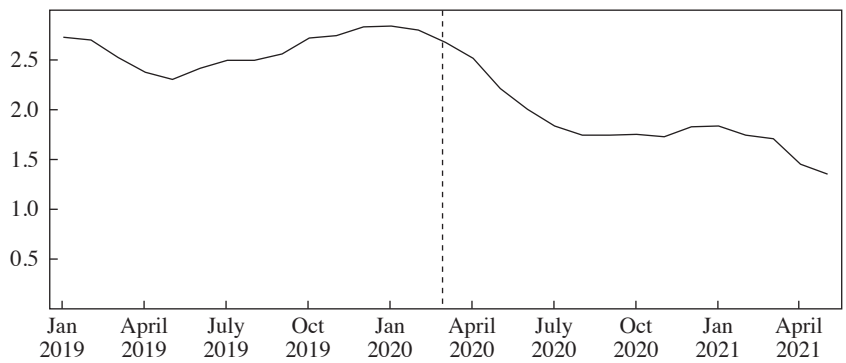
Panel A: Auto debt delinquency rate



Panel B: Auto debt forbearance rate



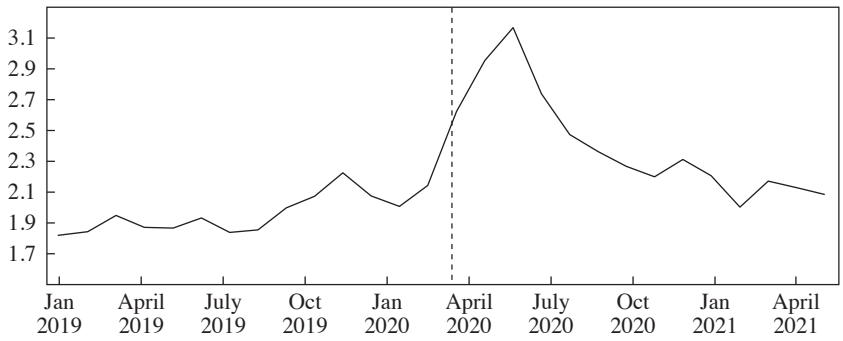
Panel C: Revolving debt delinquency rate



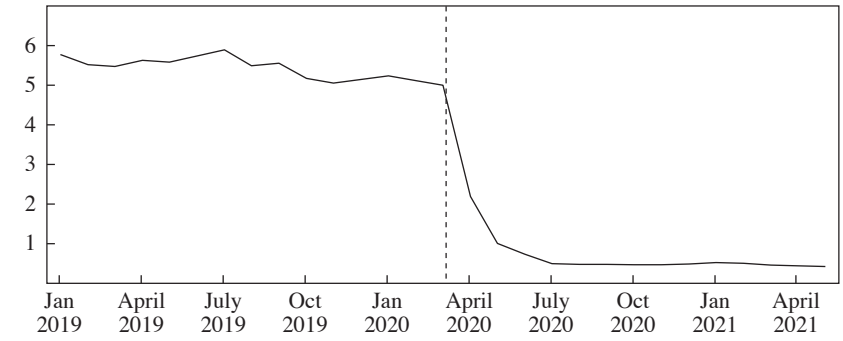
15985-03a-Cherry-4thPgs.indd 158 5/12/22 12:09 PM

Figure 2. Auto, Revolving, and Student Debt: Delinquency and Forbearance Rates (*Continued*)

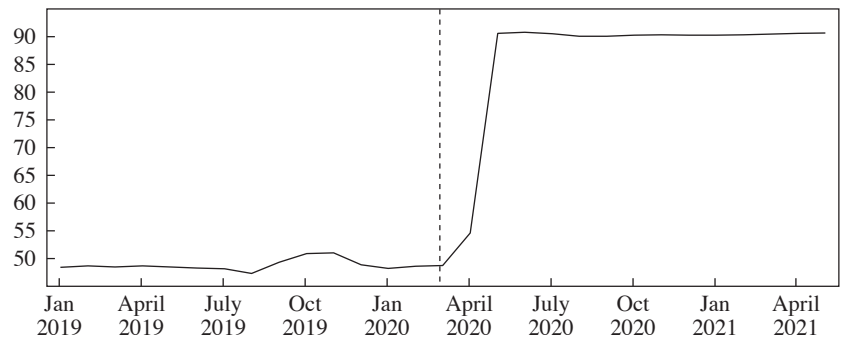
Panel D: Revolving debt forbearance rate



Panel E: Student debt delinquency rate



Panel F: Student debt forbearance rate



Source: Equifax.

Note: Delinquency rates are thirty days past due or worse. The dashed vertical line shows the declaration of the national emergency and the passage of the CARES Act in March 2020. The data from Equifax comprise a 10 percent representative sample of the US credit population.

Forbearance gives borrowers an option to stop loan payments. About 75 percent of borrowers in the case of mortgages to more than 96 percent in the case of student debt missed their scheduled payments (table 2, panel A). On the other hand, a sizable minority of borrowers in forbearance continued to make full payments on their mortgage, auto, and revolving loans. Therefore, forbearance partially acts as a credit line, allowing borrowers to draw on forbearance if needed.

Individuals in forbearance obtain substantial debt relief. Mortgage debt was the category with the largest missed payments from individuals in forbearance, with the average borrower missing about $4,254 from March 2020 until May 2021. This mainly reflects the fact that mortgage balances are much higher relative to other debt balances. Auto borrowers missed on average about $398 during the same period, while revolving debt and student debt averaged, respectively $60.70 and $312 over this period. Debt relief is therefore substantial even considering other COVID-19 relief programs. The average recipient of a stimulus check received $1,696; the average unemployment benefits going to an unemployed worker from mid-March to the end of May 2021 ranged from around $16,000 to $32,000.[12] Therefore, payments missed through mortgage debt forbearance were larger than stimulus checks on average and amounted to as much as 20 percent of the total unemployment benefits received for some individuals during the pandemic. As we discuss below, debt relief affected a substantially different population than income-based programs, such as stimulus checks.

The substantial number of individuals who used forbearance to miss their payments means that forbearance is associated with a significant amount of debt relief at the aggregate level. Our data set is a random sample of the entire US credit population, so we can scale our estimates to the aggregate level without concerns about sample representativeness (table 2, panel B). From March 2020 to May 2021, $2.37 trillion of debt entered forbearance, with most of the increase coming from mortgages ($1.4 trillion). This increase reflected 147 million loans, and 72 million borrowers, suggesting that borrowers experienced debt relief in more than one category. By May 2021, borrowers missed an overall $86.4 billion of loan payments. At this rate, forbearance would allow more than 70 million consumers to miss about $100 billion of their debt payments by the end of September 2021, when some of the key government forbearance mandates were set to expire.

12. Our calculations assume that an individual was unemployed from mid-March to the end of May 2021. We calculate these numbers by taking the average unemployment benefits in each state and adding $600 per week for the weeks from March 29 through July 25.

The $100 billion aggregate amount of debt relief was somewhat smaller than other stimulus measures and may result in significantly lower cost. About $267 billion was spent on stimulus checks, $659 billion was provided through PPP loans, and an estimated $585 billion was spent on unemployment benefits.[13] Importantly, because loans in forbearance need to be in principle repaid, the final net transfers to borrowers due to forbearance are substantially lower than temporarily missed payments. As we argue in the next section, despite a much lower cost, the targeting of temporary relief at households in distress prevented substantial household distress, and with it likely the spillover to the rest of the economy.

III.B. Aggregate Implications: The Absence of Household Debt Distress

Large economic crises such as the Great Depression and Great Recession are often accompanied by significant household debt distress, which spills over to the rest of the economy. During the Great Recession, serious delinquency rates (sixty days or more) rose from less than 2 percent in 2006 to more than 8 percent in 2010 (Piskorski and Seru 2018). Federal government programs were put in place after a significant amount of household distress had already materialized (Piskorski and Seru 2021). A large body of work shows how distressed household debt distress spilled over into aggregate house prices, employment, and consumption (e.g., Mian and Sufi 2009, 2011, 2014a). In other words, the household debt channel had significant consequences in prior economic crises.

The pandemic had a devastating effect on the real economy with the unemployment rate reaching almost 15 percent by April 2020 and a severe decline in GDP (see figure 1). Extrapolating from the strong historical association between the unemployment rate, house price changes, borrower indebtedness, and mortgage defaults, one would expect a significant rise in household debt distress during the pandemic. We measure the expected amount of mortgage delinquency using Equifax individual loan performance data from 2006 to 2017:

$$Delinquent_{ikt} = \beta_0 + \beta_1 UE_{kt} + \beta_2 \Delta HP_{kt} + \Gamma_1 X_{ikt} + \Gamma_2 \left(UE_{kt} \times X_{ikt}\right) + \Gamma_3 \left(\Delta HP_{kt} \times X_{ikt}\right) + \epsilon_{ikt} \quad (1)$$

13. US Department of Labor, "Monthly Program and Financial Data," https://oui.doleta.gov/unemploy/claimssum.asp.

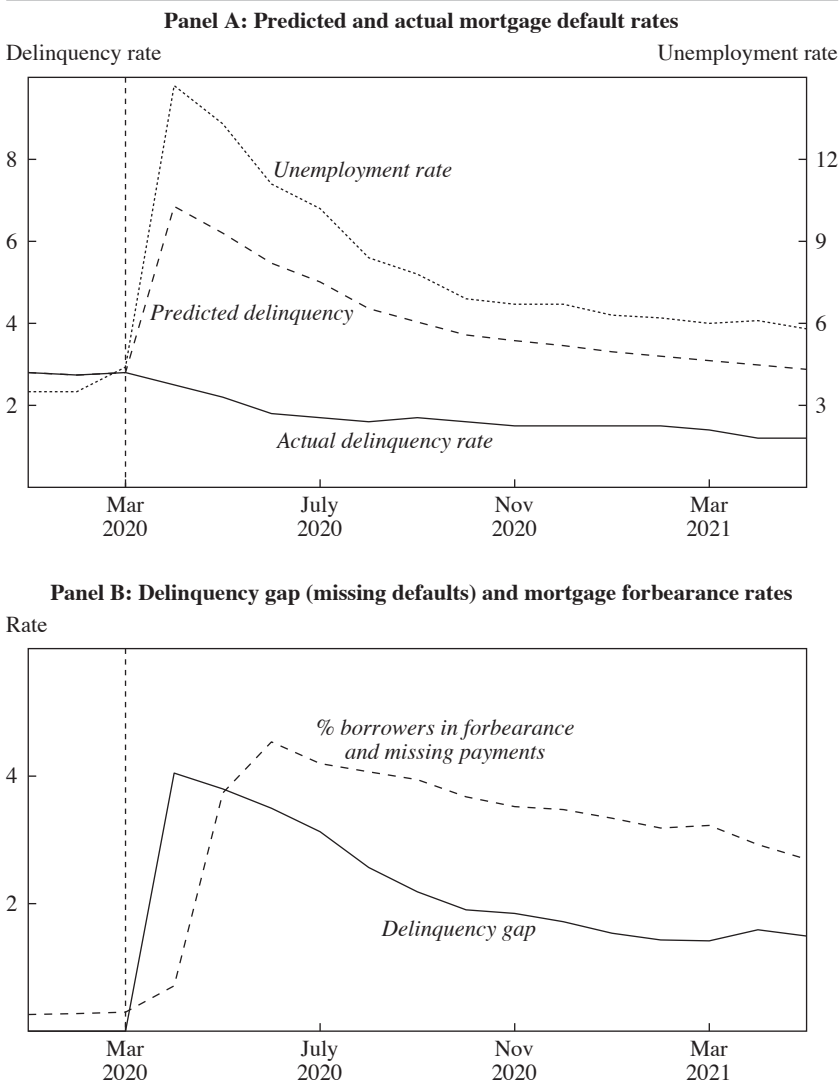
$Delinquent_{ikt}$ is an indicator for whether a mortgage loan i is thirty or more days delinquent in month t. The value UE_{kt} is the state-level unemployment rate assigned to borrower i's zip code k in month t. The values ΔHP_{kt} are the changes in house price in borrower i's zip code k in month t. The value X_{ikt} is a set of borrower and loan characteristics, including credit score, loan-to-value ratio, debt-to-income ratio, and their squared terms, which measure borrower indebtedness (home equity). Individual-level data allow us to account for changes in the composition of borrowers and their risk profile over time as well as changes in regional economic condition. The model shows a strong association of mortgage default with unemployment and changes in house prices interacted with borrower indebtedness and performs well in matching the historical aggregate delinquency rate patterns (online appendix figure A4).

The expected delinquency rate from the model would have reached the peak rate of 6.85 percent in April 2020, gradually decreasing in the following quarters but staying above 3 percent throughout the pandemic (figure 3, panel A). Instead, the *actual* mortgage delinquency rates declined during the pandemic and have remained low and steady and below 2 percent since May 2020. We call the difference between the expected (counterfactual) and realized delinquency the *delinquency gap* or *missing default rate.* The delinquency gap has averaged 3 percentage points from April 2020 to May 2021, with the highest gap of about 4.5 percentage points at the beginning of the pandemic. This delinquency gap implies about 1.5 to 2.5 million (at its peak) missing borrower defaults during the pandemic. Since defaults and foreclosures lower house prices (Campbell, Giglio, and Pathak 2011; Mian, Sufi, and Trebbi 2015; Gupta 2019), we speculate that missing defaults explain, at least in part, why the pandemic has not resulted in significant house price declines, which would have resulted in additional delinquencies, increasing the gap.[14]

The timing and level of forbearance suggest that the extent of forbearance was large enough to generate the delinquency gap. On average, for every ten missing delinquencies there are twenty mortgage borrowers in forbearance, of which about fourteen miss payments. This suggests that some borrowers may be using forbearance, even though historical data would suggest they would not default. This may reflect the fact that forbearance does not carry the usual potential costs of missing payments (e.g., lowering credit scores and risk of foreclosure). Alternatively, they may be

14. See Melzer (2017) and Diamond, Guran, and Tan (2020) for evidence regarding other effects of foreclosures.

Figure 3. Counterfactual Mortgage Defaults, Actual Defaults, and Forbearance Rates



Source: Equifax; US Bureau of Labor Statistics.

Note: Panel A compares the model-predicted mortgage delinquency rate and the actual delinquency rate along with the unemployment rate. The model is estimated using monthly loan-level data from 2006 to 2017. The dashed vertical line shows the declaration of national emergency due to COVID-19 and the passage of the CARES Act. Forbearance and missed payment rates are calculated according to Equifax's standard procedure for identifying loans in forbearance. Panel B shows the delinquency gap (prevented defaults rate), the difference between predicted mortgage delinquency rate and the actual delinquency rate and the percentage of borrowers who are in mortgage forbearance and who miss making mortgage payments. The quarterly unemployment rates are peak values in a respective quarter. The data from Equifax comprise a 10 percent representative sample of the US credit population.

using forbearance to increase their consumption during the pandemic. The pattern of forbearance over time closely matches the pattern of missing delinquencies (figure 3, panel B). The response to the pandemic differs from the Great Recession in its rapid and intensive surge in private and government debt relief. The relief materialized within weeks of the pandemic, before household distress could be fully realized (online appendix figure A3). These results suggest that debt relief has helped avoid a significant amount of household debt distress, potentially explaining why the standard household debt channel was largely absent during the COVID-19 pandemic. We validate this view in subsequent sections.

IV. Who Are the Recipients of Forbearance?

Debt forbearance has provided significant financial relief to US households throughout the initial stage of the pandemic. Ultimately, the impact of a given policy and its cost-effectiveness crucially depend on whether relief flows to those affected by shocks and in need of relief.

IV.A. Debt Forbearance across Creditworthiness and Income

We document that forbearance policies provided relief to borrowers who are less creditworthy and more likely to be liquidity constrained, even if they have higher income and are ineligible for other income-based relief programs. This take-up of debt relief may explain why forbearance has such a large impact on household distress despite a smaller aggregate expenditure than other programs.

Table 1 shows that borrowers in forbearance are more likely to be less creditworthy, have lower income, and are younger. The average Vantage credit scores are nearly 60 points lower for individuals in forbearance compared to the overall population. Borrowers in forbearance have higher average balances on all debt types, have lower average estimated annual incomes by about $6,000, and are younger by about nine years. These differences were not caused by the impact of the COVID-19 pandemic but are also present when we look at these differences in January 2020. We replicate these patterns specifically for mortgage borrowers in table A2 in the online appendix. Mortgage borrowers in forbearance were also much more likely to be delinquent on all debt types—they were over two times as likely to be delinquent on other loans, consistent with these borrowers facing credit constraints.

Forbearance rates decline in individuals' creditworthiness, income, and age (online appendix figure A6). For income and age, we create four groups

based on the 25th, 50th, and 75th percentiles of these variables. For credit scores, we simply use the four credit groups ("Low," "Fair," "Good," and "Excellent") defined by Equifax based on VantageScore. The impact of credit score is especially striking. An individual with a low credit score is more than four times more likely to be in mortgage forbearance than an individual with an excellent score. We show a decrease in forbearance across credit scores for all loan categories. The incidence of forbearance also decreases with income. The differences in forbearance are smallest for student loans, likely because all federally held student loans were automatically placed into forbearance by the CARES Act, which resulted in less-targeted, blanket financial relief across household characteristics.

Households can choose whether to miss payments upon obtaining forbearance. We find a small decline in the probability of borrowers choosing to miss payments across income, creditworthiness, and age (online appendix figure A7). These results suggest that borrower differences in creditworthiness predict whether they want to obtain the forbearance credit line and, to a much smaller extent, who chooses to use it.

Online appendix figure A8 shows that conditional on being in forbearance, the dollar amount of relief per borrower is much higher for higher-income borrowers in mortgage, auto, and student loans. For mortgages, for example, individuals in the highest income group missed over $1,000 more than individuals in the lowest income group. This is mainly because higher-income borrowers have higher loan balances and hence higher associated loan payments. The exception is the smallest debt category of revolving debt where high-income borrowers receive smaller amounts of relief. Overall, higher-income individuals are less likely to obtain relief, are similarly likely to draw on it, and once they do choose to miss payments, the amounts are larger.

We next show that the patterns in online appendix figure A6 hold up more formally when we estimate the following linear regression over the period March 2020 to May 2021:

$$y_{i,t,z} = \alpha + \beta X_{i,z} + \gamma C_z + \varepsilon_{i,t,z}, \tag{2}$$

where $y_{i,t,z}$ is an indicator variable for whether individual i who lives in zip code z is in forbearance or missed their monthly payment during month t; $X_{i,z}$ is a vector of individual characteristics such as credit score, income, debt levels and debt-to-income ratio, being a small business owner indicator, the number of accounts past due, and age, all measured as of January 2020; and C_z is either a zip code fixed effect or a vector of zip code characteristics.

Table 3. Mortgage Debt Forbearance Regressions

Panel A: Forbearance and individual characteristics

	In forbearance		*Missed payment*	
	(1)	*(2)*	*(3)*	*(4)*
Mortgage balance	0.800***	0.734***	–0.463***	–0.471***
	(0.0243)	(0.0282)	(0.0583)	(0.0702)
Auto loan balance	0.277***	0.257***	0.306***	0.293***
	(0.0119)	(0.0125)	(0.0455)	(0.0579)
Revolving debt balance	0.167***	0.184***	1.010***	0.949***
	(0.0117)	(0.0143)	(0.0395)	(0.0508)
Student loan balance	0.192***	0.195***	0.497***	0.458***
	(0.00940)	(0.0120)	(0.0376)	(0.0497)
Small business owner	0.270***	0.259***	0.476***	0.491***
	(0.00781)	(0.00968)	(0.0447)	(0.0583)
Credit score	–2.779***	–2.717***	–3.721***	–3.765***
	(0.0137)	(0.0174)	(0.0585)	(0.0777)
Income	–0.110***	–0.0720***	–0.418***	–0.334***
	(0.0168)	(0.0189)	(0.0704)	(0.0874)
Number of accounts past due in January	1.243***	1.222***	–0.208***	–0.222***
	(0.0174)	(0.0225)	(0.0319)	(0.0416)
Debt-to-income ratio	0.235***	0.310***	–0.0864	0.0203
	(0.0536)	(0.0584)	(0.0735)	(0.0590)
Age	–0.428***	–0.383***	–2.168***	–2.143***
	(0.00766)	(0.00942)	(0.0595)	(0.0778)
Zip code controls	No	Yes	No	Yes
Zip code fixed effects	Yes	No	Yes	No
Mean of dependent variable	4.622	4.433	72.91	72.73
N	67,150,585	40,375,126	3,101,291	1,789,780
Adjusted R^2	0.041	0.035	0.037	0.014

Consistent with online appendix figure A6, lower credit scores are associated with a higher probability of forbearance for all debt types. Lower income is associated with higher rates of forbearance for mortgages and student loans (table 3 for mortgages, and table A4 in the online appendix for other types of debt). Higher mortgage, auto loan, student loan, or revolving debt balances are strongly correlated with the probability of being in mortgage forbearance. We also find that other proxies for prepandemic creditworthiness predict forbearance: mortgage forbearance rates are higher for individuals with higher debt-to-income ratios and larger numbers of accounts past due as of January 2020. Finally, forbearance rates are also higher among small business owners, a group of borrowers that may have been disproportionately affected by pandemic-related restrictions.

15985-03a-Cherry-4thPgs.indd 166 5/12/22 12:09 PM

Table 3. Mortgage Debt Forbearance Regressions (*Continued*)

Panel B: Mortgage forbearance and rate

	Forbearance rate	*Scheduled payments missed (%)*
College educated (%)	–0.230***	–0.0949*
	(0.0316)	(0.0373)
Black (%)	0.526***	0.378***
	(0.0294)	(0.0429)
Hispanic (%)	0.662***	0.528***
	(0.0269)	(0.0321)
Prepandemic unemployment rate	0.157***	0.165***
	(0.0369)	(0.0406)
Median age	0.120***	0.173***
	(0.0278)	(0.0350)
Median income	0.163***	–0.0690
	(0.0381)	(0.0708)
Average debt-to-income ratio	0.235***	0.843***
	(0.0586)	(0.196)
Average house price in January	0.0308	0.111**
	(0.0271)	(0.0341)
Mean of dependent variable	4.055	2.960
N	18,078	18,078
R^2	0.0893	0.0796

Sources: Equifax; Zillow; US Census Bureau American Community Survey Five-Year Estimates.

Note: Panel A reports the results from a regression of whether an individual is in forbearance or whether an individual missed their payment while in forbearance each month on various borrower and zip code level variables. Months include March 2020 to May 2021. Columns 1 and 2 are estimated from all individuals with first mortgages; columns 3 and 4 are restricted to individuals who are in forbearance on their first mortgages. Columns 1 and 3 contain individual-level characteristics and zip code fixed effects; columns 2 and 4 contain individual level controls and zip code level controls. Zip code level controls include zip code racial composition, educational attainment, unemployment rate, median age, median income, average combined loan-to-value ratio, average debt-to-income ratio, and median house price. Regression inputs are scaled by standard deviation. Individual controls are taken as of January 2020. Standard errors are clustered at the individual level. Panel B shows the results from regressions of the average monthly forbearance rate and the percent of scheduled payments missed in each zip code on various zip code level characteristics. Average monthly forbearance rates are from March 2020 to May 2021 and the percentage of scheduled payments missed from individuals in forbearance is calculated over the same time period. Analysis is restricted to zip codes with at least ten loans. Regression inputs are scaled by standard deviation.

* = .10; ** = .05; *** = .01

We next reduce our sample to only individuals in forbearance on their mortgage accounts. The dependent variable in equation (2) is whether an individual in forbearance missed their payment. We find that lower-credit-worthiness borrowers are also more likely to draw on the forbearance credit line conditional on obtaining it for all loan categories.

There are two main takeaways from this evidence. First, we find that overall debt forbearance rates on consumer debt are much higher for less

creditworthy and more financially constrained borrowers. This differentiates debt forbearance actions from policy programs like the stimulus checks that target individuals based on their income, regardless of their actual financial conditions.

Second, while lower-income borrowers have much higher forbearance rates, conditional on being in debt forbearance, individuals with higher prepandemic incomes received by far the largest dollar amount of debt relief per individual (online appendix figure A8). This largely reflects much higher debt balances and scheduled loan payments of higher-income borrowers compared to lower-income ones. To illustrate the aggregate implications of this observation we quantify the aggregate dollar amount of financial relief due debt forbearance that flowed to borrowers with above-median income in our data ($37,000). Seventy-two percent ($22 billion) of missed mortgage payments can be accounted for by borrowers with above-median income during the March 2020–May 2021 period (table A3 in the online appendix). Similarly, 49 percent ($2.8 billion) of missed auto payments come from borrowers with above-median incomes, as did 47 percent ($2.2 billion) of missed revolving payments. The percentage of student loan missed payments from above-median individuals is much lower at just 34 percent ($15 billion).

Borrowers across different debt categories have very different levels of income. For example, mortgage borrowers in our sample have a median income of $52,000, which is much higher than the median in the overall population of consumers. Even when we define median income within each product category in panel B of table A3 in the online appendix, we find that higher-income borrowers received over 46 percent ($14 billion) of the total amount of financial relief due to mortgage forbearance. Similarly, we find that student loan borrowers with above-median income ($22,000) account for over 64 percent ($12 billion) of student loan missed payments. High-income revolving debt and auto loan borrowers account for a lower percentage of missed payments, with above-median income auto loan borrowers ($41,000) accounting for 39 percent ($2.2 billion) of total auto loan missed payments and above-median revolving borrowers ($37,000) accounting for 47 percent ($2.2 billion) of missed revolving debt payments. Across both definitions of median income, about 50 to 55 percent of the aggregate dollar amount of financial relief on all debt types flowed to borrowers with higher prepandemic incomes.

Overall, this evidence suggests that debt forbearance policies have mainly affected borrowers who are less creditworthy and more likely to be liquidity constrained. Forbearance rates on student debt, which was subject

15985-03a-Cherry-4thPgs.indd 168 5/12/22 12:09 PM

to the automatic forbearance mandate, are more uniform across borrower characteristics. Our findings also suggest that debt forbearance importantly complemented other policies targeting US consumers during the COVID-19 pandemic. Unlike policies based mainly on income, such as the stimulus check program, debt forbearance allowed less creditworthy borrowers with higher prepandemic incomes to obtain a significant amount of financial relief.

IV.B. Forbearance and Exposure to COVID-19 Shocks?

While our individual-level data include a rich set of outcomes, they lack information on characteristics that would more precisely proxy for whether a household was affected by COVID-19 shocks, such as infection with COVID-19, race, occupation, or employment. To better understand whether forbearance offered relief to households shocked by COVID-19, we turn to regional data. We explore how regional heterogeneity in the forbearance rates (online appendix figure A10) is related to the zip code socioeconomic characteristics through a series of regressions of the form:

$$y_r = \alpha_r + \beta X_r + \varepsilon_r, \tag{3}$$

where y_r is either the average forbearance rate or the percentage of debt payments missed due to forbearance in region r and X_r is a region r vector of socioeconomic characteristics. Averages are taken over the period from March 2020 to May 2021.

DEBT FORBEARANCE AND RACE We now investigate the intensity of forbearance across zip code racial composition. This analysis is motivated by the observation that the COVID-19 pandemic and its broader economic and health consequences have disproportionately affected minorities, especially Black Americans.

Areas with larger Black or Hispanic/Latino populations have higher mortgage, auto, revolving debt, and student loan forbearance rates (panel B of table 3, and table A5 and figure A9 in the online appendix). Mortgage forbearance rates are also higher in areas with higher debt-to-income ratio and higher prepandemic house prices. We also find that the amount of relief is higher in areas with larger Black and Hispanic/Latino populations across all categories of debt. We compute the amount of relief as payments missed due to forbearance divided by the sum of all scheduled payments in a zip code. This evidence combined with our individual-level findings suggests that an important share of recipients of debt forbearance

are less creditworthy, minority borrowers living in areas with higher house prices, where affordability and debt payment constraints are likely to be more binding.[15]

DEBT FORBEARANCE ACROSS INDUSTRIES AND OCCUPATION The pandemic has had a large impact on industries that require travel and face-to-face interaction, but a much smaller impact on industries where employees are able to easily work from home. Because we do not have individual-level data on borrowers' occupations, we instead investigate whether regions more exposed to industries adversely affected by the pandemic have higher forbearance rates. Zip codes with larger percentages of the population employed in agriculture, education, health, construction, and manufacturing have lower mortgage forbearance rates (table A6 in the online appendix). These occupations are typically considered essential and thus likely continued operations throughout the COVID-19 lockdowns. We also find that areas with more workers able to work from home prior to COVID-19 have lower debt forbearance rates, while regions with larger numbers of service and sales jobs have higher rates of forbearance. We find similar results for auto loan, revolving debt, and student loans in table A6 in the online appendix.

These findings highlight that both forbearance rates and missed payments are higher in regions with a larger presence of industries and occupations hit the hardest by the pandemic. Regions that have larger exposure to industries that involve travel or face-to-face interactions have higher forbearance rates, while regions with larger concentrations of essential industries or industries that can continue operations remotely have lower forbearance rates.

DEBT FORBEARANCE AND COVID-19 IMPACT Forbearance policies appear to benefit regions with high concentrations of jobs disrupted by the pandemic. We find similar results when we more directly measure which regions were most affected by COVID-19—both in terms of infection rates and in severity of lockdowns (figure A11 in the online appendix shows geographical heterogeneity in COVID-19-related characteristics). We reestimate equation (3) by regressing county forbearance rates on county-level characteristics capturing the severity of the COVID-19 impact. Counties with higher average COVID-19 case rates have higher mortgage forbearance rates (table 4 and online appendix figure A12) as well as higher forbearance rates on other types of debt (table A7 in the online appendix). These

15. See An and others (2021), who also document that lower-income and minority borrowers have received debt forbearance at higher rates.

15985-03a-Cherry-4thPgs.indd 170 5/12/22 12:09 PM

Table 4. Unemployment Claims, COVID-19 Infections, and Other Stimulus

	Forbearance rate	*Scheduled payments missed (%)*
Total unemployment insurance claims (check rate)	0.542***	0.495***
	(0.0525)	(0.0571)
Total COVID-19 case rate	0.216***	0.140*
	(0.0539)	(0.0586)
Change in credit/debit spending	0.0533	0.0673
	(0.0521)	(0.0567)
Change in time spent at workplace	–0.352***	–0.201*
	(0.0879)	(0.0956)
Families receiving full stimulus check (%)	–0.107	–0.218*
	(0.0855)	(0.0929)
Workers in at-risk industry	0.282**	0.318**
	(0.0940)	(0.102)
Amount of PPP received normalized by number of small businesses	–0.292	–0.400*
	(0.156)	(0.170)
Mean of dependent variable	3.841	2.884
N	364	364
R^2	0.3540	0.2790

Sources: Equifax; Opportunity Insights Tracker; US Census Bureau American Community Survey Five-Year Estimates.

Note: This table shows the results from regressions of the average monthly mortgage forbearance rate and the percent of scheduled payments missed in each zip code on various county-level characteristics. Averages are taken over the period from March 2020 to May 2021 and percentage of scheduled payments missed by individuals in forbearance is calculated over the same time period. Analysis is restricted to counties with at least ten loans. Regression inputs are scaled by standard deviation.

* = .10; ** = .05; *** = .01

counties also experience the largest amount of relief (missed payments). We find similar results when measuring the severity of disruption with county unemployment insurance claim rates and percentages of workers in at-risk industries, where at-risk industries are those at most risk of having operations disrupted by the pandemic.[16] On the other hand, we find that counties with greater declines in time spent at the workplace relative to January 2020 have lower forbearance rates. Thus, counties affected more by COVID-19 infections and restrictions experienced higher rates of debt forbearance.

Taken together, our individual and regional analysis suggests that debt forbearance has reached its intended target: financially vulnerable borrowers living in regions that experienced the highest COVID-19 infection rates and the greatest deterioration in their economic conditions.

16. We define at-risk industries according to criteria described by Muro, Maxim, and Whiton (2020).

DEBT FORBEARANCE AND ABSENCE OF HOUSEHOLD DISTRESS ACROSS BORROWERS AND REGIONS Using aggregate data in section III.B, we illustrate that the extent and timing of forbearance is closely related to the delinquency gap, suggesting that forbearance relief was responsible for the absence of household debt distress. The cross-sectional evidence above suggests forbearance reached households and regions that were most likely affected by COVID-19 shocks. Here, we use cross-sectional variation to connect the dots and show that forbearance is highest in the regions where we observe the biggest gap between the expected and actual defaults. In other words, forbearance can explain the absence of household distress in the cross section. Recall that the delinquency gap is computed as an aggregated predicted delinquency across individual borrowers after accounting for local fundamentals such as unemployment and house prices, as well as borrower-level indebtedness such as loan-to-value ratio interacted with local conditions, which should proxy for the fact that some borrowers are more likely to be underwater on their debt than others.

The delinquency gap is largest among borrowers who also received the most forbearance (online appendix figure A13).[17] The delinquency gap and forbearance rates decline in credit scores. Forbearance rates in the highest credit score bin are much higher than the delinquency gap, suggesting that a sizeable share of more creditworthy borrowers have been using forbearance as a liquidity facility or credit line and would not have been in default in the absence of forbearance. We also find larger missing default rates in regions with more minorities (online appendix figure A13, panel b). Regions with more than 20 percent of the population classified as minorities have 3.7 percentage points fewer delinquencies than predicted by fundamentals, which is about 60 percent higher than regions with the smallest minority shares. The delinquency gap is also largest in regions with more households below the poverty level and with lower median income. The lowest income regions have about 1.5 percentage points more missing delinquencies than the highest income regions. Likewise, the delinquency gap in the poorest regions is about 1.8 percentage points larger (about twice as high in relative terms) than that in the richest regions. In all panels, regions with larger missing defaults also receive more forbearance, indicating that the forbearance take-up was highest among borrowers whose expected debt distress was the largest.

17. Figure A14 shows a close overlap between regions with higher shares of missing defaults (panel a) and areas with higher shares of borrowers in forbearance who miss their payments (panel b).

Formally, we explore how the regional heterogeneity in the delinquency gap is related to the zip code socioeconomic characteristics through a series of regressions of the form:

$$y_r = \alpha_r + \beta X_r + \varepsilon_r \tag{4}$$

where y_r is delinquency gap in region r, and X_r is a vector of socioeconomic characteristics in region r. Averages are taken over the period from March 2020 to May 2021.

Regional socioeconomic characteristics that predict forbearance also predict missing delinquencies. Panel a of table A8 in the online appendix shows that areas with low credit scores, larger minority shares, more households below poverty level, and smaller shares of households with college degrees have larger delinquency gaps. Similarly, occupations that are correlated with forbearance are also correlated with missing delinquencies: those that were most affected by the pandemic induced shutdowns. Panel b of table A8 in the online appendix shows that zip codes with larger percentages of the population employed in education, health, agriculture, finance, construction, and manufacturing have smaller delinquency gaps. These occupations likely continued operations throughout the pandemic lockdowns and thus experienced less unemployment. On the other hand, zip codes with larger percentages employed in service occupations had larger delinquency gaps. This is reasonable since service industries were forced to either cease or dramatically reduce operations throughout the pandemic. We also find that regions with more workers able to work from home prior to the pandemic had smaller delinquency gaps. Thus, the cross-sectional evidence confirms our aggregate conclusion that forbearance was to some degree responsible for the absence of household distress.

DEBT FORBEARANCE AND OTHER COVID-19 RELIEF PROGRAMS We conclude this section by investigating the relation between debt forbearance and other stimulus measures to provide suggestive evidence on the extent to which debt forbearance complements these other policies at the region (state) level. Very broadly, the take-up of forbearance in dollar terms is positively related to unemployment benefits and PPP loans. On the other hand, states with higher amounts of stimulus check money per individual obtain relatively less relief through forbearance policies (figure 4, panels A–C).

Figure 4 shows that forbearance and other government programs in the aggregate are positively related to missing defaults. However, the comparison of panel D and E indicates that one dollar of forbearance is much more strongly related to missing defaults compared to other stimulus measures.

Figure 4. Debt Forbearance, Missing Defaults, and Other Stimulus Measures

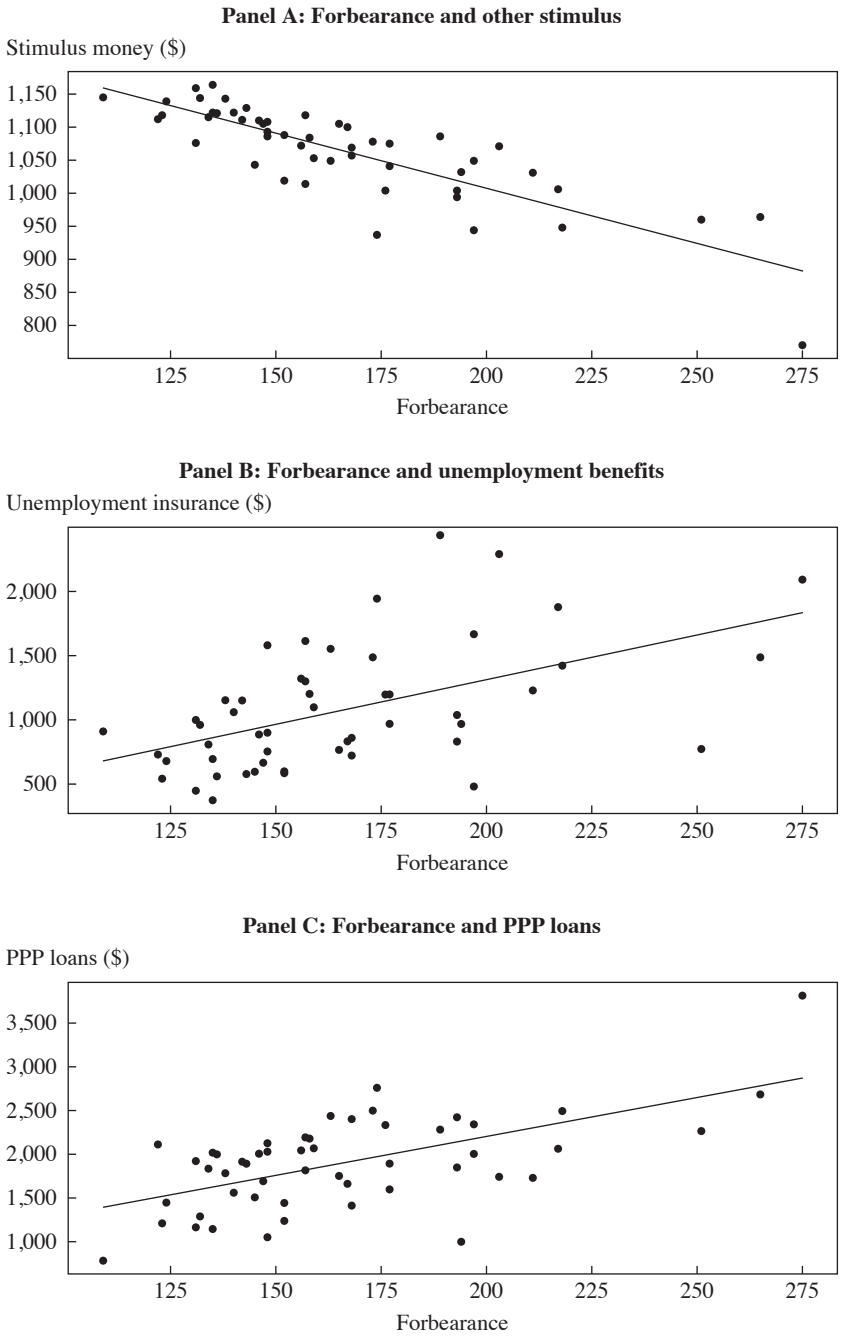
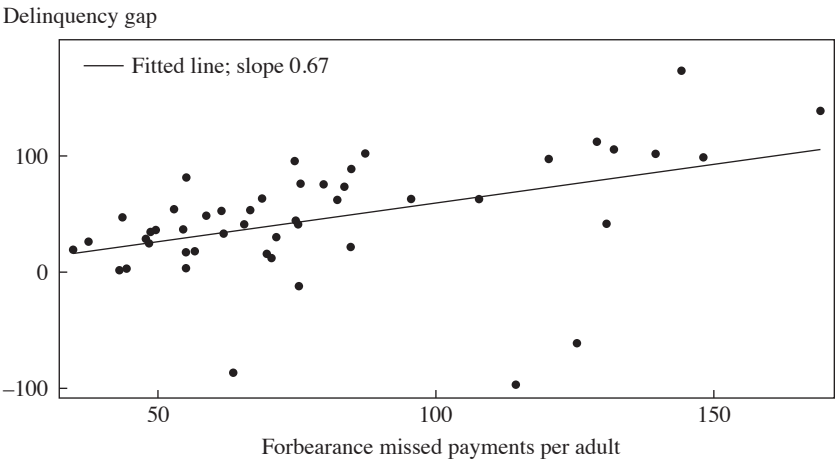
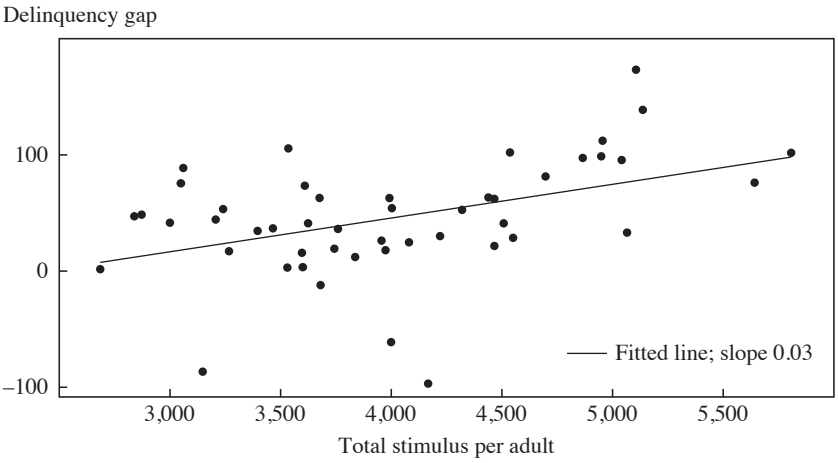


Figure 4. Debt Forbearance, Missing Defaults, and Other Stimulus Measures (*Continued*)

Panel D: Forbearance and missing defaults



Panel E: Other stimulus and missing defaults



Sources: Equifax; Internal Revenue Service; Department of Labor; Small Business Administration; and authors' calculations.

Note: This figure plots the total amount of money missed due to forbearance in a state normalized by the number of adults against the total amount of money received from stimulus checks, PPP loans, and unemployment insurance normalized by the number of adults at the state level. Panel A plots the amount missed through forbearance against the amount received from stimulus checks, panel B plots the amount missed through forbearance against the amount received from unemployment benefits, panel C plots the amount missed through forbearance against the amount received from PPP loans. Amounts are calculated from the months from March 2020 to May 2012. Panel D plots the dollar amount of missing delinquencies per borrower—the difference between model-predicted delinquency and actual delinquency multiplied by January scheduled payments—against forbearance missed payments per adult. Panel E plots missing delinquencies against the total amount of other government stimulus per adult.

For every $100 in missing payments through forbearance, we observe about $70 of missing delinquencies (slope 0.67). Every $1,000 stimulus is associated with about $30 of missing delinquency payments (slope 0.03). These associations suggest that, per dollar of stimulus, forbearance is much better targeted in terms of preventing defaults compared to other stimulus measures. Figure A15 in the online appendix shows the associations between each of these programs and missing defaults. Unemployment benefits are most closely related to dollar amount of missing delinquencies, followed by PPP loans, whereas stimulus checks do not show a positive association. This is not surprising, since stimulus checks mainly target individuals with lower incomes—that is, those individuals who are more likely to be renters or living in less expensive homes and thus are less exposed to forbearance related relief (in dollar terms).

Overall, these findings are in line with our individual-level evidence, which suggests that debt forbearance may have importantly complemented other stimulus measures. Debt forbearance provided significant financial relief to individuals and regions with higher prepandemic incomes who were not directly targeted by the income-based programs such as stimulus checks but who were nevertheless credit constrained and affected by COVID-19. There are also other factors that might have contributed to the missing household debt distress channel. For example, home equity was historically high prior to the pandemic. While our missing delinquency model captures home equity changes through the interaction of the loan-to-value ratio and house price changes, such delinquencies may not lead to foreclosures. Moreover, given how well capitalized the banking system was going into the crisis, and the Federal Reserve's aggressive stance in purchasing mortgage-backed securities and lowering interest rates (Milstein and Wessel 2021), the forced sales may not have depressed house prices or available lending. Both factors could have prevented the downward spiral through feedback of house price decline to the real economy.[18] Another factor on this front is the relaxed refinancing requirement, as a considerable amount of funds were transferred to households refinancing their mortgages. Golding and others (2020) estimate more than $100 billion transfer through this form, which is as an additional source of support during the pandemic.

18. According to Federal Reserve Bank of New York Research and Statistics Group, about half of the 1 percent increase in housing prices relative to the same period in 2019 in the months of April through August 2020 was due to a decline in sales that could be attributed to forbearance and to the lack of sales that would have resulted without forbearance (Anenberg and Scharlemann 2021).

15985-03a-Cherry-4thPgs.indd 176 5/12/22 12:09 PM

V. Government Mandates for Debt Forbearance

We conclude our analysis by studying the significant role of the private sector in extending debt relief. Financial institutions provided substantial debt forbearance in segments (auto loans and revolving debt) outside explicit government mandates as well as in mortgage segments, which were not covered by the government mandate (see figure 2). In aggregate, more than 20 percent of financial relief due to forbearance was provided outside of the government mandates. This evidence suggests that a considerable amount of debt forbearance would have existed even in the absence of explicit government mandates embedded in the CARES Act. We next exploit the variation in government mandates across different types of loans to better understand the role of two forces in determining the extent and incidence of debt relief: self-selection by borrowers and implicit forbearance subsidies by government programs.

V.A. Debt Forbearance Mandates: The Role of Self-Selection

There are generally two steps in determining which borrowers obtain debt relief. First, the borrower must request relief from the lender—self-selection. Second, the lender must agree to provide relief. As we discuss in section I.A, the CARES Act guaranteed forbearance for all government-backed mortgages, accounting for about two-thirds of outstanding mortgage debt for borrowers who requested it. Borrowers did not have to request forbearance on federally held student loans, which were automatically placed in administrative forbearance. On the other hand, auto loans and revolving debt was not directly affected by the CARES Act, so lenders were not required to grant forbearance on these debt types.

The take-up rate of mortgages covered by the government mandate is a priori evidence that self-selection is a first-order force in determining forbearance rates. Less than 10 percent of eligible borrowers requested relief, and of those, about three-quarters took advantage of the ability to defer payments. Because borrowers had to apply, but forbearance was mandated conditional on application, borrowers' self-selection is driving the forbearance rates for these loans.

Self-selection also seems to be the first-order drive of the distribution of forbearance across income and creditworthiness. We already documented that forbearance rates decline in income and creditworthiness for private and government loans across mortgages, part of which are covered by government mandates, as well as auto loans and revolving credit. In figure A16 in the online appendix, we narrow the comparison to mortgages

and consider how forbearance rates vary across borrower characteristics based on whether the mortgage is covered by the government mandate. The distribution of debt relief across borrower characteristics is quite comparable between government and private mortgages. The same holds for the fraction of borrowers in forbearance who miss payments (figure A17 in the online appendix).

Self-selection also seems to be responsible for the correlation between COVID-19 shocks and forbearance. We reestimate regressions from table 4 for mortgages stratified by whether they are eligible for government mandates (table A9 in the online appendix). Forbearance patterns are broadly similar for both types of mortgages—regions with higher unemployment insurance claims, more COVID-19 cases, and higher percentages of workers in at-risk industries have higher rates of forbearance (though some of these associations are insignificant). These results suggest that borrowers' self-selection is important in determining how relief is allocated. This self-selection might explain why the program has been effective at preventing household debt distress, despite its smaller expenditures, especially since many missed payments will eventually be repaid.

V.B. Implicit Forbearance Subsidies

Private debt relief is presumably a result of an (ex ante) mutually beneficial renegotiation. Government-mandated relief does not need to be mutually beneficial: it can result in a transfer from the lender (government) to the borrower. Such subsidies to household debt relief may very well be warranted in the presence of renegotiation frictions and aggregate spillover. Because CARES Act eligible debt differs from private debt both in its type and in the types of borrowers, it is difficult to evaluate the magnitude of subsidies.

To evaluate the importance of implicit government subsidies we exploit a size discontinuity in eligibility of mortgages for relief under the CARES Act. We first start by considering all mortgages and reestimate the specifications in table 3 with an additional control, *Mandate*, that is an indicator variable that takes on the value of one if the loan is covered by forbearance mandates under the CARES Act (Fannie Mae, Freddie Mac, FHA, and VA loans) and is zero otherwise. Panel A of table 5 shows that loans covered by the mandate are about 1.4 percentage points more likely to be in forbearance, implying about a 30 percent higher forbearance rate relative to loans not covered by the forbearance mandate. There are not economically meaningful differences in the share of borrowers missing payments between these two categories of loans.

Table 5. Government Mandates, Mortgage Forbearance, and Missing Defaults

Panel A: Forbearance and government mandates (all mortgages)

	In forbearance		*Missed payment (%)*	
	(1)	*(2)*	*(3)*	*(4)*
Mandate	1.401***	1.405***	0.430**	0.367*
	(0.0157)	(0.0191)	(0.135)	(0.173)
Mortgage balance	0.415***	0.416***	–0.0804	0.0127
	(0.0430)	(0.0412)	(0.0952)	(0.150)
Auto loan balance	0.289***	0.268***	0.132*	0.122
	(0.0132)	(0.0132)	(0.0526)	(0.0687)
Revolving debt balance	0.215***	0.232***	1.088***	1.023***
	(0.0117)	(0.0147)	(0.0439)	(0.0569)
Student loan balance	0.236***	0.242***	0.494***	0.469***
	(0.00955)	(0.0123)	(0.0395)	(0.0515)
Small business owner	0.299***	0.292***	0.486***	0.526***
	(0.00834)	(0.0104)	(0.0501)	(0.0659)
Credit score	–2.671***	–2.602***	–3.879***	–3.999***
	(0.0136)	(0.0170)	(0.0632)	(0.0843)
Income	–0.138***	–0.114***	–0.155	–0.0712
	(0.0183)	(0.0194)	(0.0831)	(0.110)
Number of accounts past due in January	1.122***	1.106***	–0.132***	–0.182***
	(0.0155)	(0.0205)	(0.0339)	(0.0445)
Debt-to-income ratio	0.188***	0.248***	0.149	0.280***
	(0.0383)	(0.0439)	(0.0903)	(0.0628)
Age	–0.558***	–0.510***	–1.799***	–1.724***
	(0.00877)	(0.0102)	(0.0638)	(0.0846)
Zip code controls	No	Yes	No	Yes
Zip code fixed effects	Yes	No	Yes	No
Mean of dependent variable	4.564	4.366	70.44	70.19
N	50,101,202	30,108,659	2,284,682	1,314,435
Adjusted R^2	0.039	0.033	0.037	0.014

(*continued on next page*)

To conduct a tighter comparison, we exploit a size discontinuity in eligibility of mortgages for relief under the CARES Act. When a loan exceeds the limits set by the Federal Housing Finance Agency, it cannot be purchased, guaranteed, or securitized by GSEs and is thus not covered by the mandate.[19] Panel A of figure 5 shows that forbearance rates are approximately 1.6 percentage points lower to the right of the conforming loan limit, which corresponds to loans not covered by the mandate. Panel B of figure 5 shows that prior to the pandemic, forbearance rates were slightly

19. These origination loan balance limits were $417,000 for a single-family home in most regions and $625,000 in some regions through most of the last decade reaching $510,400 and $765,600, respectively, by 2020 after their progressive increases during 2017–2020.

Table 5. Government Mandates, Mortgage Forbearance, and Missing Defaults (*Continued*)

Panel B: Forbearance rates and missing defaults around the conforming loan limit

	Window ± 5%			
	Forbearance rate		*Missing defaults*	
	(1)	*(2)*	*(3)*	*(4)*
Conforming	1.762***	1.722***	0.729***	0.565***
	(0.155)	(0.180)	(0.101)	(0.119)
Mean of dependent variable	5.428	5.227	2.552	2.341
Zip code controls	No	Yes	No	Yes
Zip code fixed effects	Yes	No	Yes	No
N	858,238	524,673	555,692	340,236
Adjusted R^2	0.110	0.036	0.37	0.29
	Window ± 10%			
	Forbearance rate		*Missing defaults*	
	(5)	*(6)*	*(7)*	*(8)*
Conforming	1.667***	1.639***	0.676***	0.604***
	(0.115)	(0.136)	(0.0755)	(0.0919)
Mean of dependent variable	5.358	5.173	2.556	2.385
Zip code controls	No	Yes	No	Yes
Zip code fixed effects	Yes	No	Yes	No
N	1,466,553	896,192	945,697	577,945
Adjusted R^2	0.089	0.034	0.34	0.30

Sources: Equifax; Zillow; US Census Bureau American Community Survey Five-Year Estimates.

Notes: Panel A reports the results from a regression of whether an individual is in forbearance or missed their payment while in forbearance each month on various borrower and zip code level variables. Mandate is an indicator variable that takes on the value of one if the loan is covered by forbearance mandates under the CARES Act (Fannie Mae, Freddie Mac, FHA, and VA loans) and zero otherwise. Columns 1 and 2 are estimated for all individuals with first mortgages; columns 3 and 4 are restricted to individuals who are in forbearance on their first mortgages. Columns 1 and 3 contain individual level characteristics and zip code fixed effects; columns 2 and 4 contain individual level controls and zip code level controls. Standard errors are clustered at the individual level.

Panel B of this table reports the results from a regression of whether an individual is in forbearance (columns 1, 2, 5, and 6) on various borrower and zip code level variables, including an indicator for whether the mortgage is a jumbo loan. Columns 3, 4, 7, and 8 show the corresponding results for a regression of missing defaults rate—the difference between the model-predicted delinquency probability and the actual delinquency status on various borrower and zip code level variables, including an indicator for whether the mortgage is a jumbo loan. Odd columns contain zip code level fixed effects and even columns include zip code level controls. Analysis is restricted to mortgages with balances that fall within 95–105 percent of the conforming loan limit (90–110 percent in columns 5–8). Jumbo loans are loans that fall above 100 percent of the conforming loan limit and are not covered by forbearance mandates under the CARES Act. Conforming loans are Fannie Mae and Freddie Mac mortgages that fall below 100 percent of the conforming loan limit. Standard errors are clustered at the loan level. Borrower level controls include student debt balance, auto debt balance, revolving debt balance, age, income, credit score, debt-to-income (DTI) ratio, number of accounts past due, and small business owner.

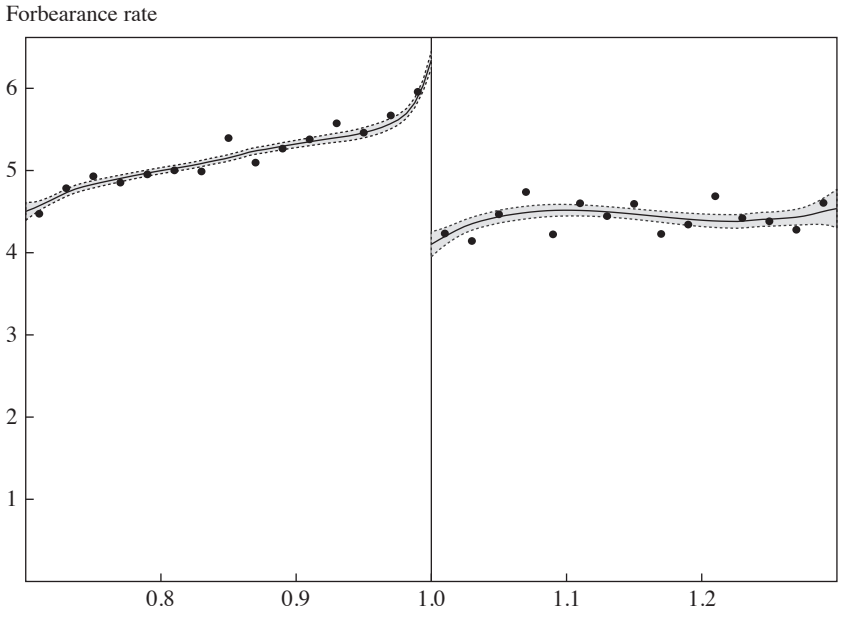
For both panels months include March 2020 to May 2021. Zip code level controls include zip code racial composition, educational attainment, unemployment rate, median age, median income, average combined loan-to-value ratio, average DTI, and median house price. Regression inputs are scaled by standard deviation. Individual (panel A) and borrower (panel B) control variables are taken as of January 2020.

* = .10; ** = .05; *** = .01

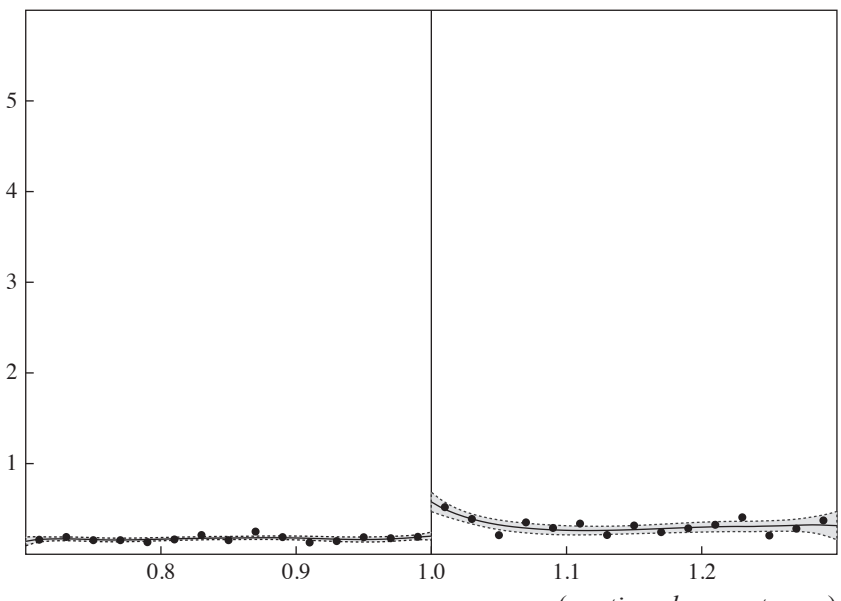
15985-03a-Cherry-4thPgs.indd 180 5/12/22 12:09 PM

Figure 5. Government Mandates: Mortgage Forbearance and Defaults around the Conforming Loan Limit

Panel A: Forbearance rate during COVID-19



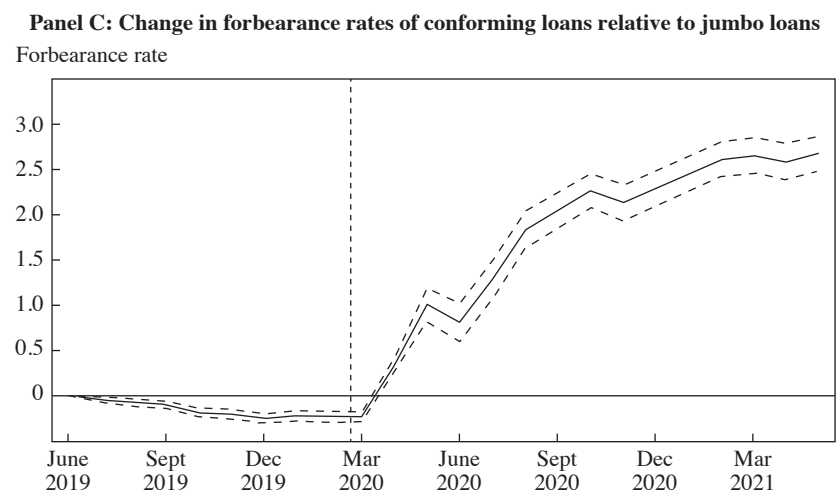
Panel B: Forbearance rate pre-COVID-19



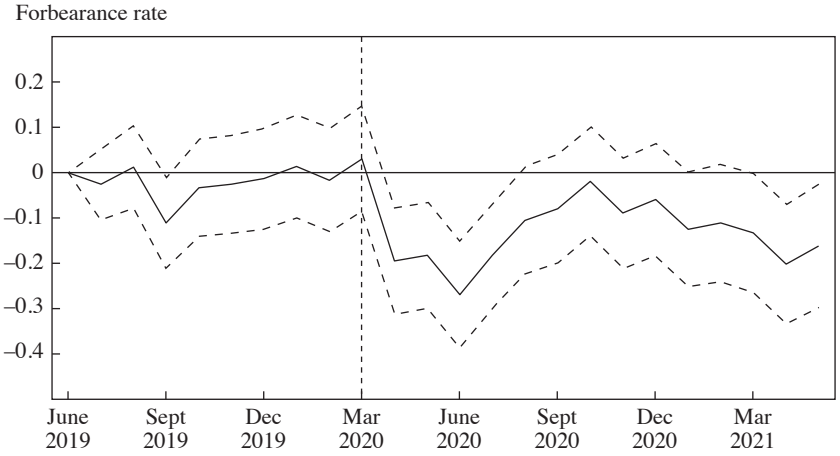
(*continued on next page*)

Figure 5. Government Mandates: Mortgage Forbearance and Defaults around the Conforming Loan Limit (*Continued*)

Panel C: Change in forbearance rates of conforming loans relative to jumbo loans



Panel D: Change in delinquency rates of conforming loans relative to jumbo loans



Source: Equifax.

Note: We restrict our analysis to mortgages with balances that fall within 70–130 percent of the conforming loan limit. Panel A shows results during the COVID-19 period (March 2020 to May 2021). Dots represent forbearance rates in 1 percent bins by the loan's percentage of the conforming loan limit. Third-degree polynomials are fit on either side of the conforming loan limit. Shaded bands show the 95 percent confidence interval. Panel B shows the same plot during the prepandemic period (January 2020 to February 2020). Jumbo loans are loans that fall above 100 percent of the conforming loan limit (solid line) and are not covered by forbearance mandates under the CARES Act. Conforming loans are loans that fall below 100 percent of the conforming loan limit and that carry the government guarantees (e.g., through Fannie Mae or Freddie Mac). Panel C shows the estimated differential changes in forbearance rates on conforming loans relative to jumbo loans along with 95 percent confidence intervals from June 2019 to May 2021. Panel D shows the corresponding results for the delinquency rate. The dashed vertical line shows the declaration of national emergency due to COVID-19 and the passage of the CARES Act. The data from Equifax comprise a 10 percent representative sample of the US credit population.

15985-03a-Cherry-4thPgs.indd 182 5/12/22 12:09 PM

higher for loans not covered by the mandate, although the difference is economically small. Panel B of table 5 confirms these results and shows that accounting for borrower and loan characteristics, mandated loans to the left-hand side of the balance limit have forbearance rates that are about 1.6–1.7 percentage points higher.

We next formally estimate the amount of government loans that obtained relief above what the private sector would have provided. We exploit the size discontinuity within a difference-in-differences design. We consider only mortgages with balances that fall within 70 percent to 130 percent of the conforming loan limit and restrict ourselves to mortgages owned by the government below the loan limit. Loans of these sizes should be relatively comparable, although those with balances below the conforming limit are covered by government forbearance mandates. We estimate:

$$Forbearance_{i,t,z} = \alpha + \beta_1 Conf_{i,z} + \beta_2 Pandemic_t + \beta_3 Pandemic_t * Conf_{i,z} + \gamma X_{i,z} + \theta Z_z + \varepsilon_{i,t,z}, \quad (5)$$

where $Forbearance_{i,t,z}$ is an indicator for whether loan i located in zip code z is in forbearance during month t. $Conf_{i,z}$ (conforming) is an indicator for whether the loan is government insured and eligible for government forbearance, $Pandemic_t$ is an indicator that takes on the value of zero prior to March 2020 and one from March 2020 to May 2021, $X_{i,z}$ is a vector of borrower-level characteristics, and Z_z is either a vector of zip code controls or fixed effects. The 1.65–1.71 percentage point coefficient on the interaction term $Pandemic_t * Conf_{i,z}$ implies that, during the pandemic, mortgages covered by the government-mandated loans were about 37 percent more likely to be in forbearance relative to ineligible mortgages (columns 1–2 of table A10 in the online appendix). Panel C of figure 5 shows the pattern of these differences over time in our sample period. The divergence between the groups begins in April–May 2020 and has continued to increase over the period. This result indicates that the mandatory approval of loan forbearance indeed increased debt relief beyond what might have been the outcome of private renegotiation between the lenders.

Taken together these results suggest that implicit government debt relief subsidies embedded in the CARES Act mandates increase the rate of forbearance by about a third. Hence, about a quarter of forbearance provided on loans covered by the government mandates is subsidized, and the rest is provided to borrowers who would have been eligible under a private

benchmark. Moreover, we find that forbearance was distributed in a similar manner across borrower characteristics in both groups (figure A16 in the online appendix), borrowers in both groups had similar rates of missing payments (figure A17 in the online appendix), and exits from forbearance are similar across the two groups. This suggests that the private sector could have increased forbearance rates toward those observed on loans covered by the mandates without sacrificing much of forbearance efficiency, at least in the short run.

Lastly, we use the discontinuity to further validate that debt forbearance can account for a substantial portion of prevented defaults during the pandemic, complementing the aggregate and cross-sectional evidence. We relate the above forbearance gap around the conforming limit to missing defaults computed in section III.B. We conduct the same regression discontinuity analyses replacing forbearance with delinquency gap as an outcome variable. Accounting for borrower and loan characteristics, the loans covered by the mandates have a delinquency gap about 0.7 percentage points higher on average (table 5, panel B). Comparing these estimates to those of the forbearance rate indicates that two mortgage forbearances are associated with one missing delinquency, the same ratio as in the aggregate data (section III.B). Finally, consistent with this observation, panel D of figure 5 (and table A10 in the online appendix) shows that after the pandemic started we see a relative decline in the delinquency rates of conforming loans relative to jumbo loans.[20]

Overall, these results suggest that the government mandates prevented additional delinquencies and validate that forbearance can account for a substantial reduction of delinquencies. The estimates are likely a lower bound on the effect of mandates. First, government mandates could have increased private forbearance supply. For example, standards set by the CARES Act might have not only provided servicers simple rules for the private sector to follow but also imposed reputational concerns on servicers who did not supply forbearance for loans not covered by the mandates. Second, when predicting delinquencies, we rely on actual house price movement during the pandemic. Without government relief programs, the realized house prices would likely be lower. Stabilizing house prices could

20. Notably, figure A18 in the online appendix shows the delinquency rates prior to the pandemic around the threshold. Prior to the pandemic, residualized delinquency rates were similar around the threshold, suggesting that the post-COVID-19 delinquency gap is not driven by preexisting differences in the default likelihood of borrowers around the conforming loan limit threshold.

15985-03a-Cherry-4thPgs.indd 184 5/12/22 12:09 PM

have prevented further distress of private loans. On the other hand, it is possible that mandated forbearance could have crowded out additional private forbearance.

V.C. The Supply of Debt Relief and the Regulatory Regime of Servicers

Servicers who collect payments from borrowers oversee which borrowers obtain forbearance. This is the case even for mortgages covered by the government mandate. Intermediary specific factors (e.g., capacity constraints) played a significant role in implementation of debt relief during the Great Recession (Agarwal and others 2017; Fuster, Lo, and Willen 2017; Piskorski and Seru 2018, 2021). One important innovation in the implementation of forbearance in the CARES Act was that the servicer did not have to determine borrowers' eligibility for forbearance. Formally, all borrowers must do is ask. So there should be little scope for differences in the intermediary supply of forbearance under government mandates. Despite that mandate, we document differences in forbearance rates across different types of servicers.

Lenders can be divided into two broad groups based on their regulatory structure: banks and shadow banks. Shadow banks—nondepository institutions—now account for about half of loan originations in the mortgage market (Buchak and others 2018; Jiang 2020; Jiang and others 2020). We show that the regulatory type of the financial intermediary is correlated with its provision of forbearance even under the same government mandate using Fannie Mae loan-level data, which have detailed information on the identity of the financial institution that originated the loan. All Fannie Mae loans had mandated forbearance under the CARES Act.

During the initial stage of the pandemic, loans originated by banks have rates of debt forbearance about 2 percentage points higher than shadow banks, and the difference persists throughout the period (panel A of figure A19 in the online appendix). We cannot completely rule out the alternative that borrowers serviced by shadow banks have different demand for forbearance. However, in table 6, we show that even controlling for extensive borrower characteristics and locations, the 2 percentage point difference persists despite their very similar characteristics. This result strongly suggests that despite the blanket government mandate, who implements forbearance has a sizeable effect on the amount of debt relief provided.

We can also divide loans based on whether the institution that originated the loan is also the servicer. Panel b of figure A19 in the online appendix shows that loans for which the servicing was outsourced also have lower rates of forbearance. In unreported results we find that this

Table 6. Forbearance Rates: Bank versus Shadow Bank Loans

	(1)	*(2)*	*(3)*
Pandemic	7.17***	7.05***	7.03***
	(0.16)	(0.15)	(0.15)
Shadow bank	0.02***	–0.71***	–0.76***
	(0.00)	(0.02)	(0.03)
Pandemic × shadow bank	−1.57***	−1.60***	−1.59***
	(0.05)	(0.05)	(0.05)
Borrower controls	No	Yes	Yes
Zip code fixed effects	No	No	Yes
R^2	0.03	0.05	0.05
N	48,896,508	48,085,901	48,085,901

Source: Fannie Mae Loan Level Performance Data.

Note: This table presents loan-level regression results of forbearance rate by bank and shadow bank. The dependent variable indicates whether the loan is in forbearance. "Pandemic" indicates whether the reporting period is after March 2020. Column 1 contains no borrower controls or zip code fixed effects, column 2 contains borrower controls but no zip code fixed effects, and column 3 contains both borrower controls and zip code fixed effects. The sample covers all loans sold to Fannie Mae since 2000 and that were still active as of January 2020. Since the GSEs do not disclose identity of servicers that service less than 1 percent of total unpaid outstanding balance, we remove loans without identified servicers. The observations are at monthly level from January 2020 to June 2020. Standard errors are clustered by zip code.

*** = .01

effect is mainly driven by bank-originated loans. One potential mechanism is that loan originators have more information about borrowers than outsourced servicers, so they can more easily tell which borrowers really require relief. This would imply that even in government-mandated relief, servicers try to evaluate whether forbearance should be granted, which would be at odds with the formal mandate. Overall, the evidence suggests that intermediary specific factors played a meaningful role in implementation of the government mandates and may have affected the overall level of debt relief provided. The regulatory framework of the servicers as well as potential access to information may have played a role in determining the supply response.

VI. Unwinding Debt Forbearance

Our results suggest that debt forbearance played an important role in muting the household debt distress channel during the pandemic by allowing millions of borrowers to postpone their debt repayments. Forbearance is designed to bridge temporary liquidity shocks faced by households—deferred payments need to be repaid. An important policy question is thus how forbearance will be unwound after it expires. We first analyze the

Table 7. Mortgage Debt Forbearance Exits

Panel A: Mortgage debt forbearance exit types

	Number of consumers	*Consumers who exit (%)*
Owe nothing	994,950	37
Reinstatement	412,690	15.4
Paid off or refinanced	116,390	4.3
Increased monthly payments	1,254,150	45
Increased balance	679,510	24
Loan modification flag	187,430	7

Panel B: Loan characteristics and forbearance exits

	Owe nothing	*Reinstatement*	*Paid off or refinanced*	*Increased balance*	*Loan modification flag*	*Increased payments*
Mortgage balance	206,194 [201,490]	215,628 [183,476]	264,513 [216,663]	215,921 [166,697]	192,581 [147,514]	209,243 [166,009]
Missed payments	0 [0]	1,322 [8,867]	7,156 [70,677]	8,747 [31,394]	4,373 [9,446]	4,981 [31,163]
Credit score	684 [101]	680 [97]	716 [79]	640 [105]	604 [100]	674 [97]
Income	55,234 [25,154]	54,607 [24,570]	63,292 [25,738]	52,858 [22,656]	45,768 [18,151]	54,137 [23,342]
Duration of forbearance (months)	4.29 [2.91]	4.67 [3.02]	4.12 [2.51]	6.28 [3.42]	6.04 [3.92]	5.4 [3.19]

Source: Equifax.

Note: The table is computed based on more than 2.6 million consumers who exited mortgage forbearance policies as of May 2021 for whom we have information regarding the manner of their exit. Standard deviations are in parentheses.

extent and nature of exits from forbearance that had already occurred in our data up to May 2021. We then discuss various ways of unwinding the substantial outstanding debt forbearance. We focus on mortgage debt forbearance, which is associated with the largest amount of debt forbearance overhang per individual (see table 2).

About 6 million mortgage borrowers entered mortgage debt forbearance during the pandemic. Of those borrowers, more than 4 million had exited forbearance as of May 2021. The remaining 1.7 million borrowers remained in forbearance, with 1.1 million borrowers missing payments. In panel A of table 7 we begin by examining forbearance exit plans for about 2.6 million mortgage borrowers who left forbearance for whom we can observe the manner of their exit.

A substantial share of borrowers either exit forbearance with no missed payments or quickly catch up on missing payments. Almost 994,000 borrowers (37 percent of those who exited) owed nothing upon exiting. Another

412,000 borrowers (15 percent) had their mortgages reinstated upon exiting forbearance, meaning they repaid all their missed payments within two months; 116,000 borrowers (4.3 percent) paid off or refinanced their loan. Panel B of table 7 shows borrower and loan characteristics by forbearance exit plans.[21] On average, borrowers who paid off or refinanced their mortgages had relatively high mortgage balances ($264,513), high incomes ($63,292), and high credit scores (716). They were also in forbearance for a relatively short period of time, with an average forbearance duration of 4.1 months. Borrowers who owed nothing or repaid missing payments in a lump sum also had relatively high credit scores and incomes and relatively short forbearance durations. In other words, for a substantial share of exiting borrowers, forbearance provided relief from temporary shocks, similar to a line of credit.

Other borrowers either increased the mortgage balance (nearly 680,000 borrowers) or increased their monthly payments (1.25 million borrowers) upon leaving forbearance. About 187,000 individuals (7 percent) had their loans modified on exit. Borrowers who exited forbearance with an increased balance, increased monthly payments, or loan modifications had lower mortgage balances on average, as well as lower incomes and credit scores. They also stayed in forbearance for a longer amount of time, with an average forbearance duration of six months. This suggests that such borrowers may have faced more severe solvency issues which temporary forbearance alone could not overcome.

In addition, close to 20 percent of mortgage borrowers who entered forbearance continued to be in forbearance and miss payments by May 2021, suggesting that they may face more serious and persistent adverse economic and financing conditions. Table 8 suggests that exit from forbearance was faster in areas with more favorable economic conditions (see figure A20 in the online appendix for regional distribution). Regions with larger unemployment increases from February 2020 to May 2021 had larger percentages of borrowers remaining in forbearance and missing payments, as did regions with higher minority shares. Regions experiencing increases in house prices from February 2020 to May 2021 had slightly lower percentages of borrowers remaining in forbearance or missing payments. Education also appears to play a role, with zip codes with a more highly educated population experiencing fewer borrowers remaining in forbearance or missing payments, as shown in table 8.

21. Our forbearance exit statistics for GSE loans are also broadly in line with reports based on the internal GSE data (McManus and Yannopoulos 2021).

15985-03a-Cherry-4thPgs.indd 188 5/12/22 12:09 PM

Table 8. Mortgage Borrowers Remaining in Forbearance and Regional Characteristics

	(1) Remaining in forbearance (%)	*(2) Remaining in forbearance with missing payments (%)*
Unemployment change	4.038***	2.470***
	(0.228)	(0.209)
House price change	–0.182	–0.194
	(0.207)	(0.184)
Median income	–0.391	0.0287
	(0.366)	(0.305)
Median debt-to-income ratio	1.294	0.846
	(0.708)	(0.463)
Black (%)	3.719***	2.742***
	(0.242)	(0.222)
Hispanic (%)	2.636***	1.629***
	(0.249)	(0.233)
College educated (%)	–2.953***	–1.623***
	(0.294)	(0.266)
House prices in January 2020	0.187	0.0797
	(0.194)	(0.175)
Median age	0.421	0.300
	(0.257)	(0.230)
Prepandemic unemployment rate	1.281***	0.642*
	(0.299)	(0.274)
N	18,590	18,590
Adjusted R^2	0.0762	0.0378

Sources: Equifax; Zillow; US Census Bureau American Community Survey Five-Year Estimates.

Note: This table shows the results from regressions of the percentage of individuals remaining in forbearance on various zip code level characteristics. In column 1, the dependent variable is the percent of individuals in a zip code who were in forbearance for at least one month during the pandemic and remained in forbearance as of May 2021. In column 2, the dependent variable is the percentage of individuals in a zip code who were in forbearance for at least one month during the pandemic and both remained in forbearance and had missing payments as of May 2021. Regression inputs are scaled by standard deviation.

* = .10; *** = .01

Looking more broadly at all debt categories, while about 60 percent of borrowers had already exited forbearance as of May 2021, more than 40 percent of 72 million Americans who entered forbearance during the pandemic were still missing about $60 billion on their debt repayments. At this rate, by September 2021 when forbearance mandates were set to expire, we estimate that these borrowers with persistent forbearance spells will be left with a forbearance overhang of more than $70 billion in accumulated postponed repayments. This estimated forbearance debt overhang amounts to about $3,900 per individual, which is about 1.5 times their average monthly income, and more than 2.2 times for lower income borrowers. For mortgage borrowers, the largest debt category, the estimated forbearance debt overhang is about $15 billion, amounting to about $14,200 per

individual (about 3.4 times their average monthly income). Moreover, as we discussed above, these borrowers with long periods of forbearance are more likely to be in lower income, higher unemployment, and higher minority share areas.

If deferred payments due to forbearance are structured as a onetime (bullet) payment, which is due immediately after forbearance ends, a significant share of these borrowers may enter distress even if this payment is anticipated. Structuring forbearance by spreading these repayments over time would alleviate households' liquidity constraints at potentially low cost to the lenders. Depending on the degree of consumer foresight and ability to smooth their consumption, the upfront versus deferred repayments could have significantly different consequences for consumers (Eberly and Krishnamurthy 2014; Piskorski and Seru 2018). Either way, the extent of forbearance overhang and its concentration among more vulnerable borrowers and regions suggest that the unwinding of forbearance could have first-order consequences for household debt distress and, through it, for the aggregate economy (Mian and Sufi 2014a).

We discuss below a few practical solutions to unwind forbearance overhang in the mortgage market, where this overhang is the largest per individual. Most of the mortgages in forbearance, including those of the most vulnerable borrowers, are insured through the government-backed programs, allowing wide latitude in implementation. The GSEs could provide clear guidance to mortgage servicers to add these missed payments to the loan balance of borrowers. Given that the remaining loan maturity for mortgage borrowers in forbearance is about twenty-five years, this would spread out the repayment of accumulated debt over a long period of time.

To illustrate this point, we use average mortgage terms of borrowers in forbearance (see table A11 in the online appendix) and consider the case of a borrower who, upon forbearance exit, misses one year of mortgage payments amounting on average to about $12,000–15,000 per borrower (FHA versus GSE borrowers). Instead of paying these amounts up front in one shot, adding these amounts to the principal balance would imply that the borrowers would face an increase in monthly mortgage payments of about $90–120 on average—a much more manageable increase in financial burden.

In addition, the government could consider a refinancing program that would allow borrowers in forbearance with federally insured mortgages to refinance their loans while increasing the loan balance of the new loan by the accumulated amount of missed payments in forbearance. Since borrowers in forbearance face mortgage rates considerably higher than the

15985-03a-Cherry-4thPgs.indd 190 5/12/22 12:09 PM

current mortgage rates (in excess of 4 percent on their legacy loans, on average), such refinancing could lower the overall mortgage payment burden of these borrowers.[22]

VII. Conclusion

We follow a representative panel of more than 20 million US consumers and analyze the widespread debt forbearance actions during the COVID-19 pandemic. While the debt delinquency rates remained low, the rate of debt forbearance that allowed households to temporarily suspend loan repayment increased dramatically following the CARES Act, providing households with significant financial relief. Between March 2020 and May 2021, 6.3 million mortgages ($1.4 trillion), 11 million auto loans ($198 billion), 68 million student loans ($655 billion), and 62 million revolving loans ($125 billion) were in forbearance. We find that the debt relief reached its intended target, since forbearance rates are higher in regions with the highest COVID-19 infection rates and the greatest local economic deterioration. Individuals with lower credit scores, lower incomes, and higher debt balances and regions with higher shares of minorities received higher rates of forbearance. Conditional on being in forbearance, borrowers with higher prepandemic income received the largest dollar amount of relief per individual. Forbearance actions, therefore, complemented other income-based policy measures, providing more than half of aggregate forbearance relief to individuals who were ineligible for policies like stimulus checks.

Not all of the forbearance was mandated under the CARES Act. The private sector provided substantial debt forbearance in segments (auto debt, revolving debt, and jumbo mortgages) outside of explicit government mandates. Moreover, allowing borrowers a choice of whether to request debt relief might have resulted in a potentially better-targeted policy.

We find that borrowers in forbearance are left with a forbearance overhang of about $60 billion in accumulated postponed repayments. This forbearance debt overhang amounts to about $3,900 per individual, which is about 1.5 times their average monthly income, and more than 2.2 times for lower income borrowers. For mortgage borrowers, the estimated forbearance debt overhang is about $15 billion, amounting to about $14,200 per individual (about 3.4 times their average monthly income). The extent of forbearance overhang suggests that the unwinding of forbearance—which

22. Federal Reserve Bank of St. Louis, FRED, "30-Year Fixed Rate Mortgage Average in the United States," https://fred.stlouisfed.org/series/MORTGAGE30US.

could be done in several ways—could have first-order consequences for household debt distress and, through it, for the aggregate economy.

Our findings suggest that massive consumer debt forbearance actions can help explain why, unlike during the Great Recession, the standard household debt distress channel was largely absent during the first year of the COVID-19 pandemic. There are several possible reasons for quicker and more widespread implementation of debt relief during the COVID-19 pandemic relative to the Great Recession. First, the private sector and policymakers may have internalized the lessons from the Great Recession pointing to significant costs of widespread defaults and foreclosures and were more willing to provide debt relief.

Second, it is possible that the underlying adverse shock has been perceived as more transitory relative to the prior crisis, which could promote a more widespread deployment of temporary debt relief measures by the private sector. This is consistent with the consumer debt design literature that indicates that lenders should provide a certain amount of debt relief during economic downturns to limit deadweight costs of default and allow better risk-sharing between borrowers and lenders, especially if the underlying shocks are transitory. Relatedly, the COVID-19 shock is a rare aggregate exogenous shock that is largely outside of the agents' influence. This should also alleviate concerns about the moral hazard effects of debt relief on incentives to repay debt, leading to a more widespread loan renegotiation effort.

Finally, the nature of financial intermediation has significantly changed since the last crisis. Prior to the Great Recession most of the riskiest consumer debt, the so-called subprime mortgages, was not retained on the balance sheets of intermediaries and instead was sold through securitization to third parties (Mian and Sufi 2009; Mayer, Pence, and Sherlund 2009; Keys and others 2010, 2013). In contrast, the private securitization market was very limited even prior to the COVID-19 pandemic, and lenders retained the majority of jumbo (non-GSE) mortgages on their balance sheets (Buchak and others 2020). As a result, the lenders may have more economic incentives to undertake voluntary debt relief actions on these loans, even though these fall outside explicit government mandates.

ACKNOWLEDGMENTS We thank Pascal Noel, Jan Eberly, Amir Sufi, Susan Wachter, and Kairong Xiao for detailed feedback. We also thank Greg Buchak, John Cochrane, Darrell Duffie, Arvind Krishnamurthy, Jim Poterba, and seminar participants at the AREUEA-ASSA meeting, Bank of England,

Columbia Leading through Crisis seminar, Housing Finance Policy Center, NBER Real Estate and Household Meetings, Northwestern, Philadelphia Federal Reserve Bank, Stanford, and Wharton, for helpful comments. Piskorski and Seru thank the National Science Foundation Award (1628895) on "The Transmission from Households to the Real Economy: Evidence from Mortgage and Consumer Credit Markets" for financial support.

References

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D. Evanoff. 2011. "The Role of Securitization in Mortgage Renegotiation." *Journal of Financial Economics* 102, no. 3: 559–578.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, Tomasz Piskorski, and Amit Seru. 2017. "Policy Intervention in Debt Renegotiation: Evidence from the Home Affordable Modification Program." *Journal of Political Economy* 125, no. 3: 654–712.

Agarwal, Sumit, Gene Amromin, Souphala Chomsisengphet, Tim Landvoigt, Tomasz Piskorski, Amit Seru, and Vincent Yao. 2020. "Mortgage Refinancing, Consumer Spending, and Competition: Evidence from the Home Affordable Refinance Program." Working Paper 21512. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w21512.

Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel. 2018. "Do Banks Pass through Credit Expansions to Consumers Who Want to Borrow?" *Quarterly Journal of Economics*, 133, no. 1: 129–90.

An, Xudong, Larry Cordell, Liang Geng, and Keyoung Lee. 2021. "Inequality in the Time of COVID-19: Evidence from Mortgage Delinquency and Forbearance." Working Paper 21-09. Philadelphia: Federal Reserve Bank of Philadelphia. https://www.philadelphiafed.org/-/media/frbp/assets/working-papers/2021/wp21-09.pdf.

Andersen, Steffen, John Y. Campbell, Kasper Meisner Nielsen, and Tarun Ramadorai. 2020. "Sources of Inaction in Household Finance." *American Economic Review* 110, no. 10: 3184–230.

Anenberg, Elliot, and Therese Scharlemann. 2021. "The Effect of Mortgage Forbearance on House Prices during COVID-19." FEDS Notes. Washington: Board of Governors of the Federal Reserve System.

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

Auclert, Adrien, Will S. Dobbie, and Paul Goldsmith-Pinkham. 2019. "Macroeconomic Effects of Debt Relief: Consumer Bankruptcy Protections in the Great Recession." Working Paper 25685. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25685.

Baker, Scott, R. A. Farrokhnia, Steffen Meyer, Michaela Pagel, and Constantine Yannelis. 2020. "Income, Liquidity, and the Consumption Response to the 2020 Economic Stimulus Payments." Working Paper 27097. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27097.

Benmelech, Efraim, Ralf R. Meisenzahl, and Rodney Ramcharan. 2017. "The Real Effects of Liquidity during the Financial Crisis: Evidence from Automobiles." *Quarterly Journal of Economics* 132, no. 1: 317–65.

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra. 2019. "Regional Heterogeneity and the Refinancing Channel of Monetary Policy." *Quarterly Journal of Economics* 134, no. 1: 109–83.

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

Berger, David W., Konstantin Milbradt, Fabrice Tourre, and Joseph Vavra. 2021. "Mortgage Prepayment and Path-Dependent Effects of Monetary Policy." *American Economic Review* 111, no. 9: 2829–78. https://www.aeaweb.org/articles?id=10.1257/aer.20181857.

Berger, David W., Nicholas Turner, and Eric Zwick. 2020. "Stimulating Housing Markets." *Journal of Finance* 75, no. 1: 277–321.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2018. "Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks." *Journal of Financial Economics* 130, no. 3: 453–83.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2020. "Beyond the Balance Sheet Model of Banking: Implications for Bank Regulation and Monetary Policy." Working Paper 25149. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25149.

Campbell, John Y., Nuno Clara, and João Cocco. 2020. "Structuring Mortgages for Macroeconomic Stability." Working Paper 27676. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27676.

Campbell, John Y., Stefano Giglio, and Parag Pathak. 2011. "Forced Sales and House Prices." *American Economic Review* 101, no. 5: 2108–31.

Capponi, Agostino, Ruizhe Jia, and David A. Rios. 2020. "The Effect of Mortgage Forbearance on Refinancing: Evidence from the CARES Act." Working Paper. Social Science Research Network, June 4. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3618776.

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

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber. 2020. "Labor Markets during the COVID-19 Crisis: A Preliminary View." Working Paper 27017. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27017.

Corcoran, Emily Wavering, and Nicholas Haltom. 2020. "Mortgage and Student Loan Forbearance during the COVID-19 Pandemic." Federal Reserve Bank of Richmond. https://www.richmondfed.org/publications/research/coronavirus/economic_impact_covid-19_08-27-20.

Cox, Natalie, Peter Ganong, Pascal Noel, Joseph Vavra, Arlene Wong, Diana Farrell, Fiona Greig, and Erica Deadman. 2020. "Initial Impacts of the Pandemic on Consumer Behavior: Evidence from Linked Income, Spending, and Savings Data." *Brookings Papers on Economic Activity*, Summer, 35–82.

Diamond, Rebecca, Adam M. Guren, and Rose Tan. 2020. "The Effect of Foreclosures on Homeowners, Tenants, and Landlords." Working Paper 27358.

Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27358.

Di Maggio, Marco, Amir Kermani, Benjamin J. Keys, Tomasz Piskorski, Rodney Ramcharan, Amit Seru, and Vincent Yao. 2017. "Interest Rate Pass-Through: Mortgage Rates, Household Consumption and Voluntary Deleveraging." *American Economic Review* 107, no. 11: 3550–88.

Di Maggio, Marco, Amir Kermani, and Christopher J. Palmer. 2020. "How Quantitative Easing Works: Evidence on the Refinancing Channel." *Review of Economic Studies* 87, no. 3: 1498–528.

Eberly, Janice, and Arvind Krishnamurthy. 2014. "Efficient Credit Policies in a Housing Debt Crisis." *Brookings Papers on Economic Activity*, Fall, 73–119.

Eichenbaum, Martin, Sergio Rebelo, and Arlene Wong. 2019. "State Dependent Effects of Monetary Policy: The Refinancing Channel." Working Paper 25152. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25152.

Elenev, Vadim, Tim Landvoigt, and Stijn Van Nieuwerburgh. 2020. "Can the COVID Bailouts Save the Economy?" Working Paper 27207. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27207.

Favilukis, Jack, Sydney C. Ludvigson, and Stijn Van Nieuwerburgh. 2017. "The Macroeconomic Effects of Housing Wealth, Housing Finance, and Limited Risk Sharing in General Equilibrium." *Journal of Political Economy* 125, no. 1: 140–223.

Fuster, Andreas, Aurel Hizmo, Lauren Lambie-Hanson, James Vickery, and Paul S. Willen. 2021. "How Resilient Is Mortgage Credit Supply? Evidence from the COVID-19 Pandemic." Working Paper 28843. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28843.

Fuster, Andreas, Stephanie H. Lo, and Paul S. Willen. 2017. "The Time-Varying Price of Financial Intermediation in the Mortgage Market." Working Paper 23706. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w23706.

Fuster, Andreas, and Paul S. Willen. 2017. "Payment Size, Negative Equity, and Mortgage Default." *American Economic Journal: Economic Policy* 9, no. 4: 167–91.

Ganong, Peter, and Pascal Noel. 2020. "Liquidity versus Wealth in Household Debt Obligations: Evidence from Housing Policy in the Great Recession." *American Economic Review* 110, no. 10: 3100–3138.

Golding, Edward, Laurie S. Goodman, Richard Green, and Susan Wachter. 2020. "The Mortgage Market as a Stimulus Channel in the Covid-19 Crisis." White Paper. Philadelphia: Penn Institute for Urban Research. https://penniur.upenn.edu/uploads/media/4-Golding-Goodman-Green-Wachter.pdf.

Granja, João, Christos Makridis, Constantine Yannelis, and Eric Zwick. 2020. "Did the Paycheck Protection Program Hit the Target?" Working Paper 27095. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27095.

Greenwald, David. 2016. "The Mortgage Credit Channel of Macroeconomic Transmission." Working Paper. Social Science Research Network, February 21. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2735491.

Greenwald, David, Tim Landvoigt, and Stijn Van Nieuwerburgh. 2017. "Financial Fragility with SAM?" *Journal of Finance* 76, no. 2: 651–706.

Guerrieri, Veronica, Guido Lorenzoni, Ludwig Straub, and Iván Werning. 2020. "Macroeconomic Implications of COVID-19: Can Negative Supply Shocks Cause Demand Shortages?" Working Paper 26918. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26918.

Guerrieri, Veronica, and Harald Uhlig. 2016. "Housing and Credit Markets: Booms and Busts." In *Handbook of Macroeconomics, Volume 2B*, edited by John B. Taylor and Harald Uhlig. Amsterdam: North Holland.

Gupta, Arpit. 2019. "Foreclosure Contagion and the Neighborhood Spillover Effects of Mortgage Defaults." *Journal of Finance* 74, no. 5: 2249–301.

Guren, Adam M., Arvind Krishnamurthy, and Timothy J. McQuade. 2021. "Mortgage Design in an Equilibrium Model of the Housing Market." *Journal of Finance* 76, no. 1: 113–68.

Guren, Adam M., Alisdair McKay, Emi Nakamura, and Jón Steinsson. 2018. "Housing Wealth Effects: The Long View." Working Paper 24729. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24729.

Hsu, Joanne W., David A. Matsa, and Brian T. Melzer. 2018. "Unemployment Insurance as a Housing Market Stabilizer." *American Economic Review* 108, no. 1: 49–81.

Hurst, Erik, Benjamin J. Keys, Amit Seru, and Joseph Vavra. 2016. "Regional Redistribution through the US Mortgage Market." *American Economic Review* 106, no. 10: 2982–3028.

Jiang, Erica Xuewei. 2020. "Financing Competitors: Shadow Banks' Funding and Mortgage Market Competition." Working Paper. Social Science Research Network, April 14. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3556917.

Jiang, Erica Xuewei, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2020. "Banking without Deposits: Evidence from the Shadow Bank Call Reports." Working Paper 26903. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26903.

Kaplan, Greg, Kurt Mitman, and Giovanni L. Violante. 2020. "The Housing Boom and Bust: Model Meets Evidence." *Journal of Political Economy* 128, no. 9: 3285–345.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig. 2010. "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans." *Quarterly Journal of Economics* 125, no. 1: 307–62.

Keys, Benjamin J., Tomasz Piskorski, Amit Seru, and Vikrant Vig. 2013. "Mortgage Financing in the Housing Boom and Bust." In *Housing and the Financial Crisis*, edited by Edward L. Glaeser and Todd Sinai. Chicago: University of Chicago Press.

Kruger, Samuel. 2018. "The Effect of Mortgage Securitization on Foreclosure and Modification." *Journal of Financial Economics* 129, no. 3: 586–607.

Looney, Adam, and Constantine Yannelis. 2019. *The Consequences of Student Loan Credit Expansions: Evidence from Three Decades of Default Cycles*. Washington: Brookings Institution.

Lopez, Mark Hugo, Lee Rainie, and Abby Budiman. 2020. "Financial and Health Impacts of COVID-19 Vary Widely by Race and Ethnicity." Pew Research Center, May 5. https://www.pewresearch.org/fact-tank/2020/05/05/financial-and-health-impacts-of-covid-19-vary-widely-by-race-and-ethnicity/.

Maturana, Gonzalo. 2017. "When Are Modifications of Securitized Loans Beneficial to Investors?" *Review of Financial Studies* 30, no. 11: 3824–57.

Mayer, Christopher, Edward Morrison, Tomasz Piskorski, and Arpit Gupta. 2014. "Mortgage Modification and Strategic Behavior: Evidence from a Legal Settlement with Countrywide." *American Economic Review* 104, no. 9: 2830–57.

Mayer, Christopher, Karen Pence, and Shane M. Sherlund. 2009. "The Rise in Mortgage Defaults." *Journal of Economic Perspectives* 23, no. 1: 27–50.

McManus, Douglas, and Elias Yannopoulos. 2021. "COVID-19 Mortgage Forbearances: Drivers and Payments Behavior." *Journal of Structured Finance* 27, no. 2: 13–25.

Melzer, Brian T. 2017. "Mortgage Debt Overhang: Reduced Investment by Homeowners at Risk of Default." *Journal of Finance* 72, no. 2: 575–612.

Mian, Atif, and Amir Sufi. 2009. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *Quarterly Journal of Economics* 124, no. 4: 1449–96.

Mian, Atif, and Amir Sufi. 2011. "House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis." *American Economic Review* 101, no. 5: 2132–56.

Mian, Atif, and Amir Sufi. 2012. "The Effects of Fiscal Stimulus: Evidence from the 2009 Cash for Clunkers Program." *Quarterly Journal of Economics* 127, no. 3: 1107–42.

Mian, Atif, and Amir Sufi. 2014a. "What Explains the 2007–2009 Drop in Employment?" *Econometrica* 82, no. 6: 2197–223.

Mian, Atif, and Amir Sufi. 2014b. *House of Debt: How They (and You) Caused the Great Recession, and How We Can Prevent It from Happening Again*. Chicago: University of Chicago Press.

Mian, Atif, Amir Sufi, and Francesco Trebbi. 2015. "Foreclosures, House Prices, and the Real Economy." *Journal of Finance* 70, no. 6: 2587–634.

Milstein, Eric, and David Wessel. 2021. *What Did the Fed Do in Response to the COVID-10 Crisis?* Washington: Hutchins Center on Fiscal and Monetary Policy, Brookings Institution.

Mueller, Holger M., and Constantine Yannelis. 2019a. "The Rise in Student Loan Defaults." *Journal of Financial Economics* 131, no. 1: 1–19.

Mueller, Holger M., and Constantine Yannelis. 2019b. "Reducing Barriers to Enrollment in Federal Student Loan Repayment Plans: Evidence from the

Navient Field Experiment." Working Paper. https://cmepr.gmu.edu/wp-content/uploads/2019/08/MuellerYannelis.pdf.

Muro, Mark, Robert Maxim, and Jacob Whiton. 2020. "The Places a COVID-19 Recession Will Likely Hit Hardest." Blog post, March 17, Brookings, The Avenue. https://www.brookings.edu/blog/the-avenue/2020/03/17/the-places-a-covid-19-recession-will-likely-hit-hardest/.

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

Piskorski, Tomasz, and Amit Seru. 2018. "Mortgage Market Design: Lessons from the Great Recession." *Brookings Papers on Economic Activity*, Spring, 429–513.

Piskorski, Tomasz, and Amit Seru. 2021. "Debt Relief and Slow Recovery: A Decade after Lehman." *Journal of Financial Economics* 141, no. 3: 1036–59.

Piskorski, Tomasz, Amit Seru, and Vikrant Vig. 2010. "Securitization and Distressed Loan Renegotiation: Evidence from the Subprime Mortgage Crisis." *Journal of Financial Economics* 97, no. 3: 369–97.

Piskorski, Tomasz, and Alexei Tchistyi. 2010. "Optimal Mortgage Design." *Review of Financial Studies* 23, no. 8: 3098–140.

Piskorski, Tomasz, and Alexei Tchistyi. 2011. "Stochastic House Appreciation and Optimal Mortgage Lending." *Review of Financial Studies* 24, no. 5: 1407–46.

Piskorski, Tomasz, and Alexei Tchistyi. 2017. "An Equilibrium Model of Housing and Mortgage Markets with State-Contingent Lending Contracts." Working Paper 23452. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w23452.

Scharfstein, David, and Adi Sunderam. 2016. "Market Power in Mortgage Lending and the Transmission of Monetary Policy." Working Paper. https://www.hbs.edu/ris/Publication%20Files/Market%20Power%20in%20Mortgage%20Lending%20and%20the%20Transmission%20of%20Monetary%20Policy_8d6596e6-e073-4d11-83da-3ae1c6db6c28.pdf.

Wachter, Susan M., John V. Duca, and Lilit Popoyan. 2019. "Real Estate and the Great Crisis: Lessons for Macroprudential Policy." *Contemporary Economic Policy* 37, no. 1: 121–37.

Comments and Discussion

COMMENT BY

PASCAL NOEL There is normally a strong positive relationship between unemployment and mortgage default. Yet during the COVID-19 pandemic, the unemployment rate rose dramatically while mortgage defaults fell. This paper by Susan Cherry, Erica Jiang, Gregor Matvos, Tomasz Piskorski, and Amit Seru provides an explanation for this surprising pattern. They show that the novel expansion of mortgage debt forbearance initiated by the Coronavirus Aid, Relief, and Economic Security (CARES) Act can account for the missing defaults relative to what would have been expected based on economic fundamentals alone. In total, 6.3 million mortgages entered forbearance between March 2020 and May 2021, with borrowers being permitted to miss $31 billion in payments on these mortgages.

The paper extends beyond this basic conclusion in four significant ways. First, the authors show that the relationship between higher mortgage forbearance and lower defaults is not just a correlation. They present compelling evidence that increased availability of forbearance caused a reduction in defaults. Second, they quantify the substantial impact of forbearance across all consumer debt products, not just mortgages: borrowers of student loans, auto loans, and credit cards were also permitted to miss $55 billion in payments through forbearance over this time period. Third, the authors address distributional questions, showing that forbearance was relatively well targeted overall. In particular, it appears to have helped households negatively affected by the pandemic but who were unlikely to be eligible for other income-based programs. Finally, they show that who provided this forbearance mattered. Traditional banks were more likely to provide forbearance in their role as intermediaries than were shadow banks.

Overall, this paper suggests that the aggressive forbearance policy pursued during the pandemic was largely successful. The goal of my discussion

15985-03b-Cherry_Comm&GD-4thPgs.indd 200 5/12/22 12:09 PM

is to expand on this finding by comparing the (largely successful) forbearance policy implemented during the COVID-19 recession to the (less successful) debt relief policy implemented in the Great Recession. In particular, I will briefly describe the main government debt relief policy implemented during the Great Recession (the Home Affordable Modification Program, or HAMP), summarize lessons learned from a decade of research assessing HAMP's limitations, argue that the pandemic debt relief policy addressed each of HAMP's three main limitations, and conclude by considering implications for future policy.[1] My comments focus on forbearance in the mortgage market.

HAMP TWISTED THE PAYMENT SCHEDULE BY CREATING AN ENTIRELY NEW LOAN

During the Great Recession, HAMP was the main government program designed to restructure distressed mortgage debt. In total, around 1.8 million borrowers received modifications through the program.

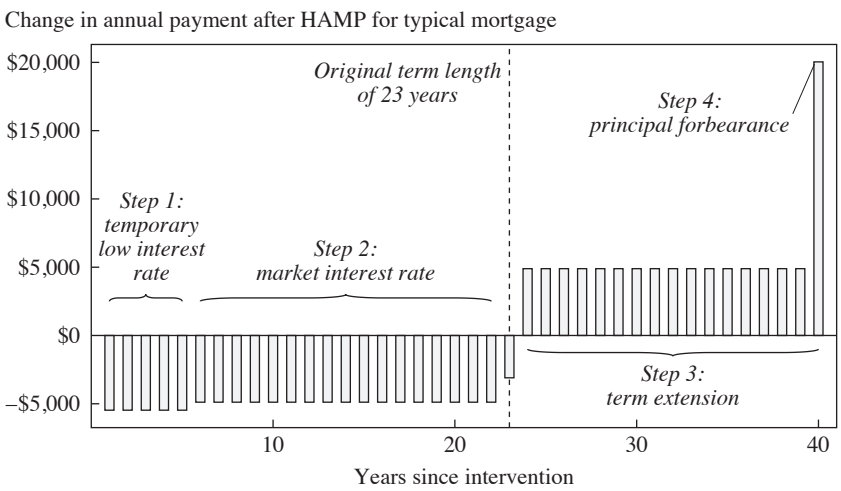
The program had three important characteristics. First, it was heavily subsidized by taxpayers. Second, it was implemented by servicers. While the government guarantees and sets standards for a significant portion of mortgage debt, for the most part it does not directly issue mortgages or interface with individual borrowers. Thus, any program to address distressed debt within the prevailing market structure needed to go through mortgage servicers, who intermediate between borrowers and lenders. Third, HAMP was designed as a modification of the original mortgage contract. Therefore, it required the execution of an entirely new contract with new terms and, crucially, new underwriting. Furthermore, this new contract varied across borrowers on a case-by-case basis.

The reason that HAMP required a new mortgage contract is that it provided payment relief to borrowers via a complicated restructuring of each mortgage's specific terms. The target was to reduce a borrower's payments to 31 percent of their income for a period of at least five years. A HAMP modification involved as many as four steps to achieve this target.

First, the interest rate was reduced down to a floor of 2 percent for a period of five years. Second, the interest rate gradually increased to the market rate prevailing at the time of the modification, essentially providing a refinance. Third, if the target was not reached after the interest rate changes, the mortgage maturity was extended out to forty years. Finally, if the target was still not reached, a portion of the unpaid principal balance was converted into a non-interest-bearing balloon payment due at the end

1. US Department of the Treasury, "Home Affordable Modification Program (HAMP)," https://home.treasury.gov/data/troubled-assets-relief-program/housing/mha/hamp.

Figure 1. Structure of Typical Great Recession Mortgage Modification



Source: Author's calculations.

Note: This figure simulates the annual cash-flow impact of the standard HAMP modification structure if it were applied to a mortgage with the average premodification characteristics of loans that received CARES forbearance. The premodification characteristics used are those for GSE loans reported in table A11 in the online appendix. The modification is calibrated to match an initial payment reduction of 35 percent, the average reduction observed during the Great Recession (Ganong and Noel 2020a, appendix table 1). This is achieved by simulating multiple modifications that followed the HAMP protocol with an initial payment reduction between 5 percent and 65 percent, to capture the impact of all four possible modification steps.

of the mortgage term. This structure meant that the exact level of the new payment varied according to each borrower (since it depended on their individual income), as did the exact terms of the new "modified" mortgage contract.

Figure 1 shows how cash flows are affected under the standard HAMP modification structure. To ease comparison between HAMP and CARES Act interventions, the figure applies the HAMP modification structure to the typical mortgage that entered a CARES Act forbearance plan.[2] Figure 1 depicts the change in annual payments under HAMP relative to the payments owed according to the premodification mortgage terms. The average HAMP modification resulted in a 35 percent payment reduction, or about $5,500 per year for the typical mortgage. As figure 1 shows, these payment

2. I use the characteristic for government-sponsored enterprises (GSEs) loans from table A11 in the online appendix. This borrower had a premodification interest rate of 4.4 percent, a principal balance of $223,000, and a remaining loan term of twenty-three years.

15985-03b-Cherry_Comm&GD-4thPgs.indd 202 5/12/22 12:09 PM

reductions remain significant throughout the original mortgage term. Past the initial mortgage term of twenty-three years, payments would have been higher due to the term extension and principal forbearance components of the modification. Thus, the four modification steps in HAMP twisted the payment schedule, with reduced payments initially offset by increased payments far in the future.

LIMITATIONS AND LESSONS LEARNED SINCE HAMP While HAMP did provide assistance to millions of borrowers, analysis since the Great Recession has pointed to three challenges that limited its impact. First, HAMP may have been on the wrong side of the trade-off between type I and type II error. In the political climate of 2009, policymakers expressed worry about providing bailouts to borrowers who didn't need or deserve government assistance. This can explain HAMP's strict documentation guidelines. Borrowers needed to provide detailed documentation of their current income, attest to their lack of savings, explain the reasons behind their current financial hardship, and fill out necessary paperwork. Furthermore, servicers needed to verify and keep track of all this paperwork and pass documents back and forth for borrower signatures and final execution. However, the strict guidelines induced by the worry about false positives (admitting undeserving applicants) led to many false negatives: it was far too difficult to obtain a modification and millions of borrowers were turned away (GAO 2012). Over time, the government recognized this challenge and responded by relaxing documentation requirements.

Second, the program required intensive effort by servicers with limited capacity, and many intermediaries were sluggish at providing assistance. For example, Agarwal and others. (2017) document substantial heterogeneity across different servicers, with a few large servicers in particular offering modifications at half the rate of others. Agarwah and others (2017) show that total modifications could have increased by 70 percent if these less active servicers renegotiated loans at the pace of their more active counterparts. Furthermore, the paper shows that these low modification rates were due to servicer-specific factors related to preexisting organization capacity, which was necessary to execute the complex HAMP modifications. As another example, Piskorski, Seru, and Vig (2010) show that ownership is also important in limiting complex renegotiations: securitized loans were renegotiated at lower rates than bank-held loans.

Third, HAMP provides drawn-out payment relief, while research since the Great Recession suggests that immediate liquidity is the key driver of mortgage defaults. As figure 1 shows, HAMP provides payment relief spread out across many years. However, a large body of literature now

points to the key role of current liquidity in driving defaults. For example, in prior research with Peter Ganong, I have found that payment reductions more than five years in the future have no impact on default, while immediate payment reductions have large impacts on default (Ganong and Noel 2020a). Other research on the strong relationship between current payments and default includes Hsu, Matsa, and Melzer (2018), Piskorski and Seru (2018), Tracy and Wright (2016), Fuster and Willen (2017), Di Maggio and others (2017), Agarwal and others (2020), Ganong and Noel (2020b), Scharlemann and Shore (2017), Ehrlich and Perry (2015), and Abel and Fuster (2021).

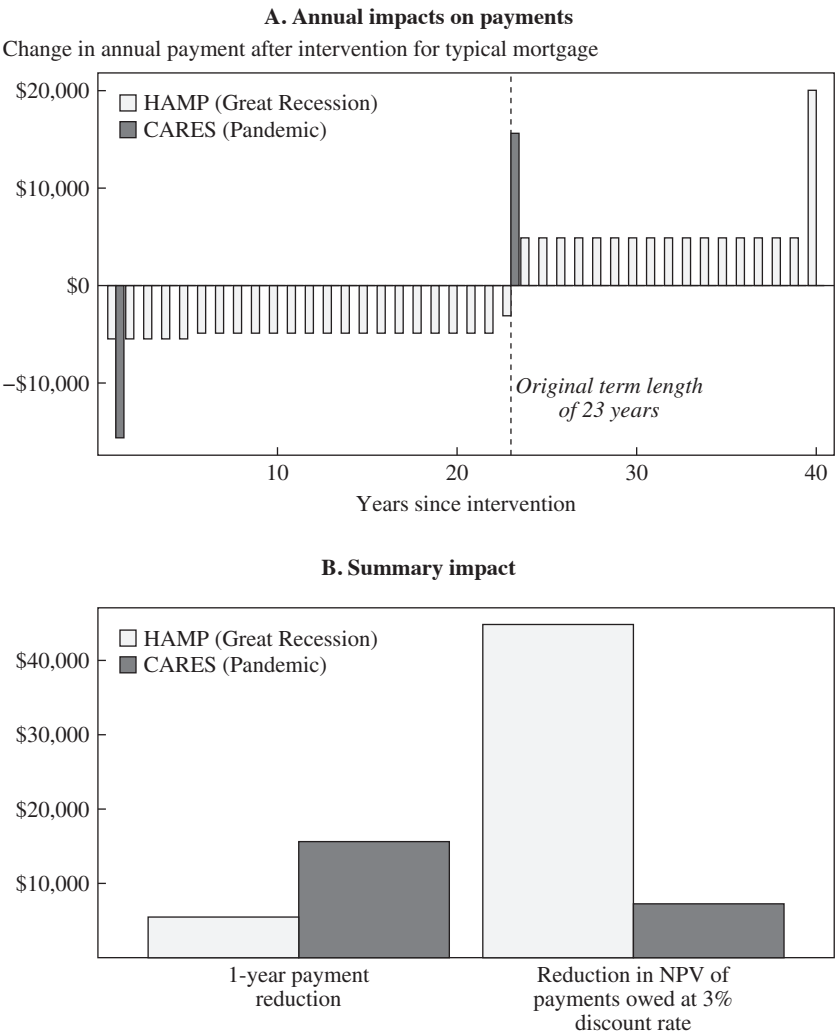
HOW CARES ACT FORBEARANCE ADDRESSES EACH OF THESE LIMITATIONS The design of mortgage forbearance as mandated through the CARES Act is significantly different than the design of HAMP modifications. One clear difference is its impact on borrower cash flows. Figure 2, panel A, compares the annual impacts on cash flows for the typical CARES Act forbearance to the impact on cash flows for the typical HAMP modification. While HAMP reduced payments by 35 percent on average, forbearance plans reduce payments by 100 percent for the duration of the plan. Mortgages covered by the CARES Act were eligible for a maximum of about eighteen months of forbearance, though figure 2 depicts the cash flow impacts for an average borrower taking up forbearance for twelve months.

Although the initial payment reduction provided by a forbearance is much deeper, it is also more temporary. After the forbearance plan ends, a range of options are available for making up the missed payments. One option emphasized by regulators and policymakers was payment deferral, whereby the missed payments are due at the end of the original mortgage term, essentially turned into a non-interest-bearing balloon payment.[3] This payment deferral option is what is modeled in figure 2. The borrower saves $15,600 during the one year of forbearance, goes back to making regular monthly payments when deferral ends, and then owes the $15,600 of missed payments at the end of the mortgage term (or refinance or sale, whichever comes first). Moving payments from the present to the future reduces the net present value (NPV) of total mortgage payments due. Figure 2, panel B, summarizes this NPV impact using an interest rate of 3 percent, which was the average market mortgage interest rate during

3. Federal Housing Finance Agency, "FHFA Announces Payment Deferral as New Repayment Option for Homeowners in COVID-19 Forbearance Plans," https://www.fhfa.gov/Media/PublicAffairs/Pages/FHFA-Announces-Payment-Deferral-as-New-Repayment-Option-for-Homeowners-in-COVID-19-Forbearance-Plans.aspx.

15985-03b-Cherry_Comm&GD-4thPgs.indd 204 5/12/22 12:09 PM

Figure 2. Financial Impact of CARES Forbearance Relative to HAMP Modifications



Source: Author's calculations.

Note: This figure simulates the impact of the typical HAMP modification and CARES forbearance structure, applied to a mortgage with the average premodification characteristics of loans that received CARES forbearance. For CARES forbearance, this figure assumes that forbearance lasts one year, and the missed payments are added to the end of the mortgage term (i.e., payment deferral). Panel A plots the difference in average annual payments for borrowers receiving each type of intervention relative to the payments owed by borrowers under their status quo unmodified mortgage contracts. Panel B summarizes the financial impacts of interventions along two dimensions: the change in the first year payment and the change in the net present value of mortgage payments owed, discounted at a 3 percent interest rate.

the CARES Act forbearance period. The $15,600 of immediate payment reduction only costs lenders $7,300 in NPV cash flows, since the borrower must eventually repay. In sharp contrast, HAMP keeps payments low for much longer, costing substantially more in NPV terms to lenders (and taxpayers, who subsidized these modifications). Figure 2, panel B, shows that the $5,500 one-year payment reduction in HAMP costs lenders $44,800 in NPV terms.

Figure 3 depicts the impact of CARES Act forbearance on cash flows under two alternative exit strategies also commonly provided to borrowers. First, borrowers who had the funds could immediately repay at the end of the forbearance plan. This amounts to a one-year interest-free loan. Second, borrowers who could afford slightly higher monthly payments could have the missed payments capitalized into the total principal balance and amortized over the rest of the initial loan term, resulting in an increase of about $120 in the monthly payment. Each of these alternative exits were even cheaper than the payment deferral option from the lender's perspective. Indeed, because the prepandemic interest rate for the typical mortgage was above the prevailing 3 percent interest rate during the pandemic, capitalization actually has a negative cost to lenders.[4]

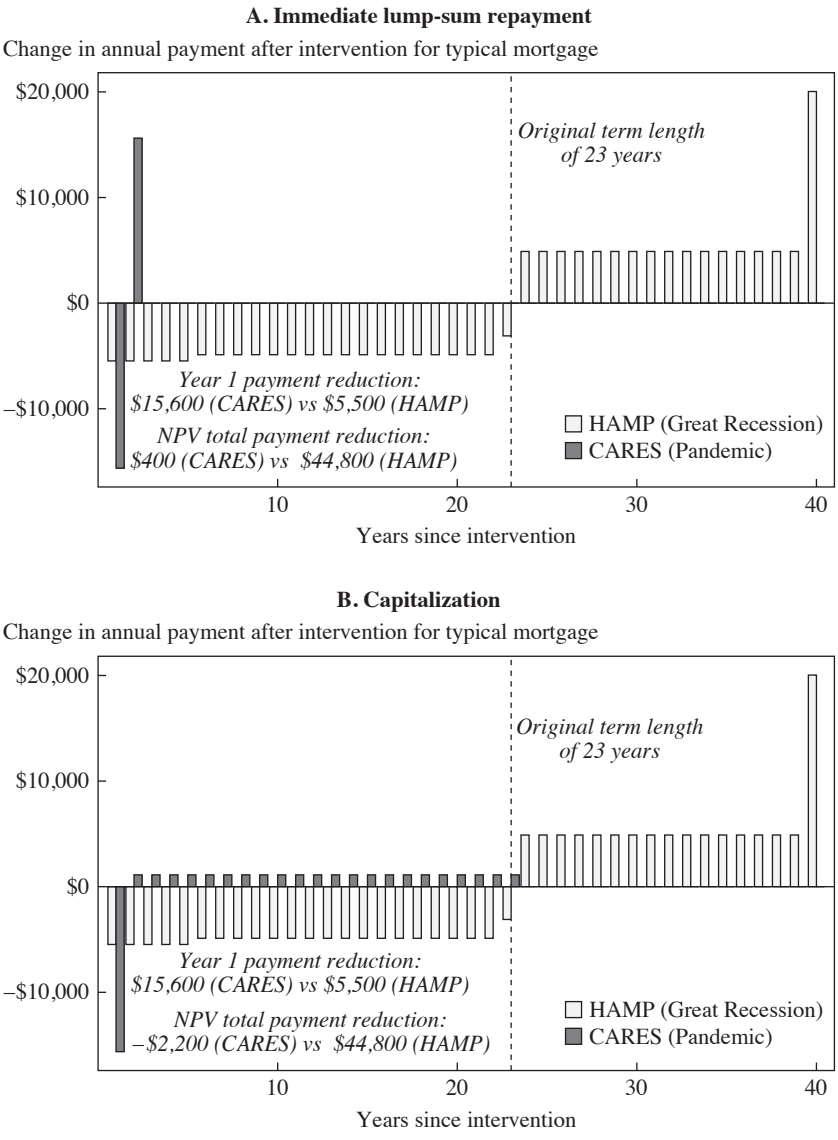
This cash flow design enabled CARES Act forbearance to address each of the three main limitations of HAMP. First, CARES Act forbearance required no documentation. Borrowers simply needed to tell their servicer that they had a pandemic-related hardship and request forbearance. This simple process avoided the issue of widespread false negatives from HAMP: essentially any borrower who wanted a forbearance could get one. The fact that CARES Act forbearance was substantially cheaper for lenders in an NPV sense helps justify this much more lax screening. If lenders aren't losing substantial money by providing this forbearance, it is less crucial for them to carefully limit entry.

Second, CARES Act forbearance is much simpler for the intermediaries to implement. There is no documentation requirement, no up-front re-underwriting, and no up-front modification to mortgage terms. Servicers simply pause payments and just need to keep track of the total dollar amount of missed payments. This design does not require the same organizational capacity as HAMP, found lacking in Agarwal and others (2017).

4. When the forborne payments are capitalized into the balance of the loan, the lenders earn the interest rate on the loan, which averages 4.4 percent in the authors' sample. This is higher than the interest rate of 3 percent that prevailed for new loans at this time. For this reason, capitalization has a negative NPV cost to lenders.

15985-03b-Cherry_Comm&GD-4thPgs.indd 206 5/12/22 12:09 PM

Figure 3. Financial Impact of Alternative CARES Forbearance Exits



Source: Author's calculations.

Note: This figure compares the financial impacts of HAMP modifications to two alternative structures for CARES forbearance. Panel A shows the impact of an immediate lump sum repayment of the forborne amount, also known as a bullet payment. Panel B shows the impact of capitalizing the missed payments into the unpaid principal balance and recalculating the monthly payment under the pre-forbearance mortgage term and interest rate.

Nevertheless, the evidence in this paper suggests that there is still room for improvement on this dimension. In particular, there appears still to be some implementation challenges with shadow banks, despite the simple design. Providing automatic forbearance to 100 percent of borrowers, as was done for student loans, would fully address this challenge, though at a much larger financial cost and, as the authors note, worse targeting.

Third, and perhaps most striking, the mix between current and future payment reductions is completely flipped between CARES Act forbearance and HAMP modifications. CARES Act forbearance provides almost three times the immediate liquidity at one-sixth the NPV cost. This maximal provision of immediate cash flow relief to borrowers is exactly what is suggested by recent research.

IMPLICATIONS FOR FUTURE POLICY This paper makes a compelling case that aggressive forbearance initiated by the CARES Act was effective at preventing a spiral in mortgage defaults. The success of this intervention shows that a policy design that reflects the lessons learned since the Great Recession can work in practice. In particular, a policy that is broadly available, simple to implement, and maximizes immediate liquidity can reach millions of borrowers quickly and prevent acute distress.

Does the success of CARES Act forbearance suggest that policymakers should implement a policy like this in a future crisis? Although there are reasons to be cautious about this conclusion, I believe the answer is yes. As I have argued, the design of CARES Act forbearance addresses each of the major shortcomings of prior interventions. And as the authors have shown, it actually worked.

Given the success, why might we be worried about generalizing this remedy to alternative environments? There are at least three reasons. First, the housing market coming into the recession was very strong. In particular, households had substantial equity. If defaults are sometimes driven by negative home equity alone, then a policy focused on borrower liquidity may not work in other situations when homeowners are more likely to be underwater. However, my own view, influenced by my recent research, is that negative cash flow shocks are necessary conditions for nearly all mortgage defaults (Ganong and Noel 2020b), so it seems likely that a liquidity-focused policy would work even in a scenario with falling prices and underwater borrowers.

Second, the pandemic was an obvious negative external shock that was outside of any individual household's control. If widespread forbearance was made available in a future crisis, would households who didn't actually need assistance abuse the program? The authors' finding that less than

10 percent of eligible borrowers actually took up the forbearance option suggests that this may not be a concern in practice.

Finally, forbearance makes the most sense when there is a temporary shock, but payments can't be reduced to zero forever in the face of a more permanent shock. However, at the onset of a crisis, it is difficult for lenders and policymakers to triage borrowers into those needing temporary versus permanent assistance. This desire to tailor policy to each borrower's specific individual circumstances is exactly what made HAMP unwieldy in practice. It seems preferable to provide maximal, but temporary, liquidity up front and then use the intervening time to triage borrowers. Indeed, the authors find that two-thirds of borrowers exited forbearance on their own, before their eligibility expired. In this circumstance, lenders and policymakers can then focus on those remaining borrowers who reveal a need for deeper assistance. Furthermore, by buying time at the onset of a crisis, the policy itself can reduce the share of borrowers likely to experience permanent shocks by improving macroeconomic conditions.

Therefore, I think the arguments in favor of implementing a policy like this again outweigh the potential concerns under many possible scenarios. I hope we never face a severe housing crisis again. But if we do, I hope policymakers recognize the success of the forbearance policy documented in this paper and consider a similar approach.

REFERENCES FOR THE NOEL COMMENT

Abel, Joshua, and Andreas Fuster. 2021. "How Do Mortgage Refinances Affect Debt, Default, and Spending? Evidence from HARP." *American Economic Journal: Macroeconomics* 13, no. 2: 254–91.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, Tomasz Piskorski, and Amit Seru. 2017. "Policy Intervention in Debt Renegotiation: Evidence from the Home Affordable Modification Program." *Journal of Political Economy* 125, no. 3: 654–712.

Agarwal, Sumit, Gene Amromin, Souphala Chomsisengphet, Tim Landvoigt, Tomasz Piskorski, Amit Seru, and Vincent Yao. 2020. "Mortgage Refinancing, Consumer Spending, and Competition: Evidence from the Home Affordable Refinancing Program." Working Paper 21512. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w21512.

Di Maggio, Marco, Amir Kermani, Benjamin J. Keys, Tomasz Piskorski, Rodney Ramcharan, Amit Seru, and Vincent Yao. 2017. "Interest Rate Pass-Through: Mortgage Rates, Household Consumption, and Voluntary Deleveraging." *American Economic Review* 107, no. 11: 3550–88.

Ehrlich, Gabriel, and Jeffrey Perry. 2015. "Do Large-Scale Refinancing Programs Reduce Mortgage Defaults? Evidence from a Regression Discontinuity Design."

Working Paper 2015-06. Washington: Congressional Budget Office. https://www.cbo.gov/publication/50871.

Fuster, Andreas, and Paul S. Willen. 2017. "Payment Size, Negative Equity, and Mortgage Default." *American Economic Journal: Economic Policy* 9, no. 4: 167–91.

Ganong, Peter, and Pascal Noel. 2020a. "Liquidity versus Wealth in Household Debt Obligations: Evidence from Housing Policy in the Great Recession." *American Economic Review* 110, no. 10: 3100–3138.

Ganong, Peter, and Pascal Noel. 2020b. "Why Do Borrowers Default on Mortgages? A New Method for Causal Attribution." Working Paper 27585. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27585.

Government Accounting Office (GAO). 2012. *Foreclosure Mitigation: Agencies Could Improve Effectiveness of Federal Efforts with Additional Data Collection and Analysis*. GAO-12-296. Washington: Author. https://www.gao.gov/products/gao-12-296.

Hsu, Joanne W., David A. Matsa, and Brian T. Melzer. 2018. "Unemployment Insurance as a Housing Market Stabilizer." *American Economic Review* 108, no. 1: 49–81.

Piskorski, Tomasz, and Amit Seru. 2018. "Mortgage Market Design: Lessons from the Great Recession." *Brookings Papers on Economic Activity*, Spring, 429–513.

Piskorski, Tomasz, Amit Seru, and Vikrant Vig. 2010. "Securitization and Distressed Loan Renegotiation: Evidence from the Subprime Mortgage Crisis." *Journal of Financial Economics* 97, no. 3: 369–97.

Scharlemann, Therese, and Stephen H. Shore. 2017. "The Effect of Mortgage Payment Size on Default and Prepayment: Evidence from HAMP Resets." Working Paper. https://www.aeaweb.org/conference/2018/preliminary/paper/SQSrbS2d.

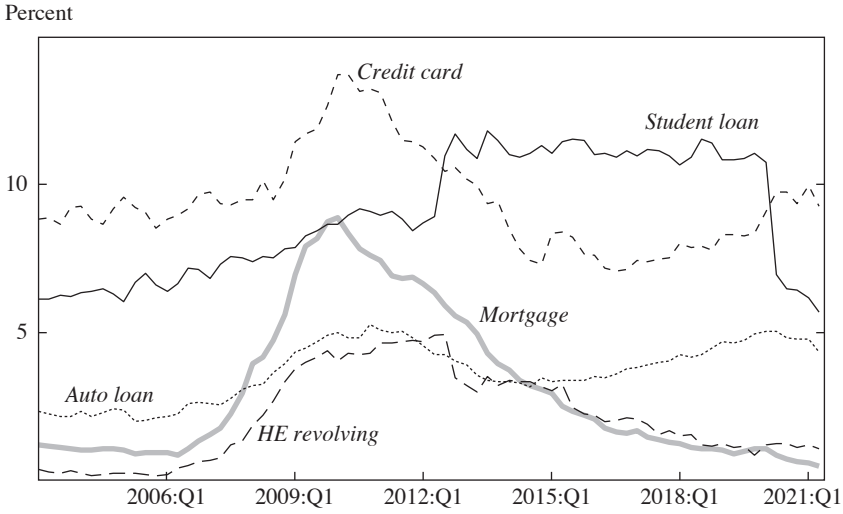
Tracy, Joseph, and Joshua Wright. 2016. "Payment Changes and Default Risk: The Impact of Refinancing on Expected Credit Losses." *Journal of Urban Economics* 93, no. C: 60–70.

COMMENT BY

SUSAN WACHTER With the onset of COVID-19 in early 2020, Congress quickly passed the Coronavirus Aid, Relief, and Economic Security (CARES) Act to help indebted households. Despite unemployment hitting 15 percent highs in April 2020, mortgage delinquencies declined with the act's implementation, as the law intended. This paper documents the public and private debt relief that the law provided and the positive outcomes for avoiding debt distress. The authors' documentation of these outcomes is an important contribution to the evaluation of debt relief assistance for policymakers and future historians.

The paper analyzes the results of debt forbearance in the aggregate and by credit type, identifying how the mechanisms of the law varied with

Figure 1. Delinquency across Debt Types



Source: New York Fed Consumer Credit Panel/Equifax.

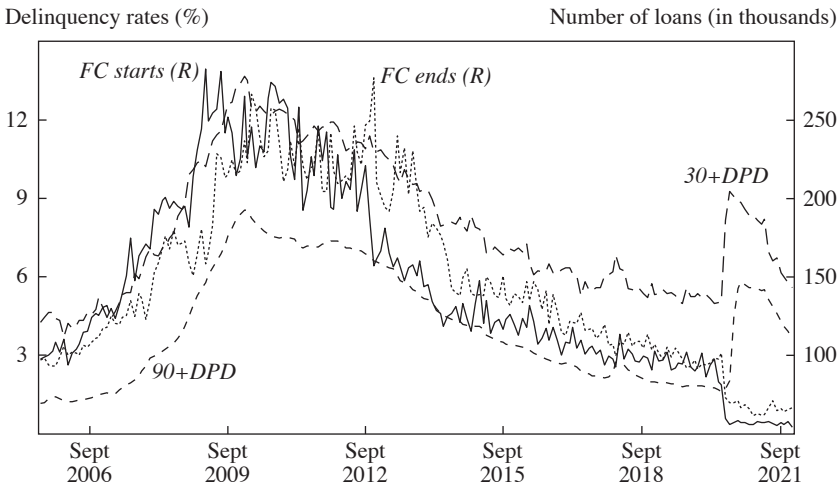
differing outcomes for the various categories of household debt, including for mortgages (both government and private), auto loans, student loans, and credit card debt. The act mandates forbearance for mortgage debt provided by government-backed entities (75 percent of mortgage providers) upon a simple request, with no documentation needed (Kim and others 2018). That is, publicly backed mortgage borrowers self-selected for assistance. The law put into place automatic relief with no request needed for student debt. In contrast, borrowers needed to negotiate debt relief with private providers of auto loans and consumer credit and for privately backed mortgages.[1]

As figure 1 indicates, student debt and mortgage debt delinquency plummeted, with student debt payments on hold and with substantial use of mortgage forbearance, while credit card and auto loan delinquencies remained elevated. As shown in figure 2, according to a report by the Philadelphia Federal Reserve Bank (RADAR 2021), which counts forborne loans as delinquent, foreclosure activity stopped abruptly in March

1. The act required that there be no reporting of missed debt payments to credit scoring agencies once forbearance is granted. Gerardi and others (forthcoming) show the importance of this in preventing deterioration in borrowers' credit ratings.

15985-03b-Cherry_Comm&GD-4thPgs.indd 211 5/12/22 12:48 PM

Figure 2. Delinquency Rates and Foreclosure Flows as of September 2021



Source: Black Knight McDash Data and Black Knight Data and Analytics, LLC. Used with permission from the Federal Reserve Bank of Philadelphia, www.philadelphiafed.org/consumer- finance/mortgage-markets/examining-resolution-of-mortgage-forbearances-and-delinquencies- october-2021.

Note: These delinquency figures reflect investor reporting and will not match reporting to the credit bureaus; the CARES Act prohibits reporting as delinquent if the mortgage was current on March 1, 2020, and is past due and in CARES Act forbearance.

2020 and fell to unprecedented low levels.[2] The authors' analysis shows that the law delivered debt relief, most importantly for the large category of mortgage debt, to intended beneficiaries.[3] In particular, the authors show that forbearance mitigated temporary liquidity problems, as about 50 percent of borrowers either did not use or repaid their forbearance requests within a few months. Self-selection of borrowers for mortgage forbearance was instrumental in providing assistance quickly and to those in immediate need. While much of the forborne debt has been repaid, the paper shows that as of September 2021, 40 percent remained outstanding and needed to be repaid going forward.

2. The report shows that even with the expiration of the federal foreclosure moratorium on July 31, 2021, foreclosure activity has not increased; however, the report notes that such activity would be likely to rise after Consumer Financial Protection Bureau safeguards end on January 1, 2022.

3. The paper demonstrates that forbearing debt collection benefited borrowers in need, particularly by regional exposure to COVID-19 and by region minority status. While individual borrower characteristics on minority status were not available to the authors, see An and others (2022) for similar results based on individual characteristics.

15985-03b-Cherry_Comm&GD-4thPgs.indd 212 5/12/22 12:09 PM

The paper also uncovers two somewhat surprising, stylized facts, discussed further below: government-backed and non-government-backed (jumbo) mortgage lenders (without mandates) provided similar levels of debt relief. And despite the same requirements across servicers of government-backed loans, nonbanks delivered less debt relief, all else equal.

A major conclusion of the paper is that debt forbearance, brought about by public and private action, prevented debt distress such as occurred during the global financial crisis. The centerpiece contribution of the paper is the measurement of how and by how much mortgage debt relief prevented mortgage debt delinquency or, in the authors' term, debt distress.

In the empirical analysis, the authors estimate missing delinquencies, that is, the mortgage delinquencies that would have been expected due to elevated unemployment but that did not occur. To calculate missing delinquencies, the authors use the historical relationship of unemployment to mortgage delinquency to predict what mortgage delinquency rates would have been in the absence of debt relief. Using data from 2007 through 2017, the authors regress mortgage delinquencies (thirty days or more) against the unemployment rate, with house price change, loan characteristics, and location factors as controls. Using parameters from the historical relationship, the authors forecast expected mortgage delinquencies from March 2020 through September 2021 and then subtract actual delinquencies from predicted delinquencies to calculate their measure of missing defaults. The authors use the Equifax Analytic Dataset, constructed to be a randomized 10 percent sample of the US population, which includes credit information on payment history as well as characteristics of households and their debt and debt payment history.

The authors estimate between 1 million and 2.5 million missing delinquencies. They also estimate that the number of forbearance requests and forbearance requests used were substantially higher than this. They therefore reasonably conclude that the number of loans requested and used in forbearance was more than sufficient to account for missing delinquencies.

The authors go on to discuss the implications of missing defaults for financial stability. If borrowers cannot make their mortgage payments, forced sales or foreclosures may follow. All else equal, the additional supply of housing on the market would produce downward pressure on housing prices. In the global financial crisis, lenders foreclosed on 8 million homes, with downward price spirals due to the additional housing supply forcing additional foreclosures (Levitin and Wachter 2020).

15985-03b-Cherry_Comm&GD-4thPgs.indd 213 5/12/22 12:09 PM

In the pandemic, a Federal Reserve Board study (Anenberg and Scharlemann 2021) shows that this effect did not occur. The study estimates that prices were higher by 0.6 percentage points in the months of April through August 2020, relative to the same period in 2019, due to this missing mechanism of forced sales. As house price growth increased on average about 1 percentage point over that period, this is not a small effect. Forbearance contributed to house price stability during this period.

From the second quarter of 2020 onward, housing prices increased at a far more rapid rate, with overall housing prices rising about 10 percent over the year. Other substantial government support, including monetary easing together with the rapid adoption of work-from-anywhere technology, caused a surge in the demand for homes. Refinancing at the now substantially lower mortgage rates lowered mortgage payments, easing forbearance exits.[4]

Preconditions also mattered to the ability to repay forborne debt. Data from the Federal Reserve of Philadelphia's RADAR group show that by June 2021, only 2 percent of households were underwater with negative equity as compared to 45 percent in 2009 (An and others 2022). The banking system was well capitalized going into the crisis, as were the government-sponsored entities (GSEs) Fannie Mae and Freddie Mac. Policies put into place in the aftermath of the global financial crisis that strengthened the mortgage system enabled borrowers to repay their forborne debt and lenders to refinance and lend more. As the authors state, it was illiquidity that was the problem, not insolvency.

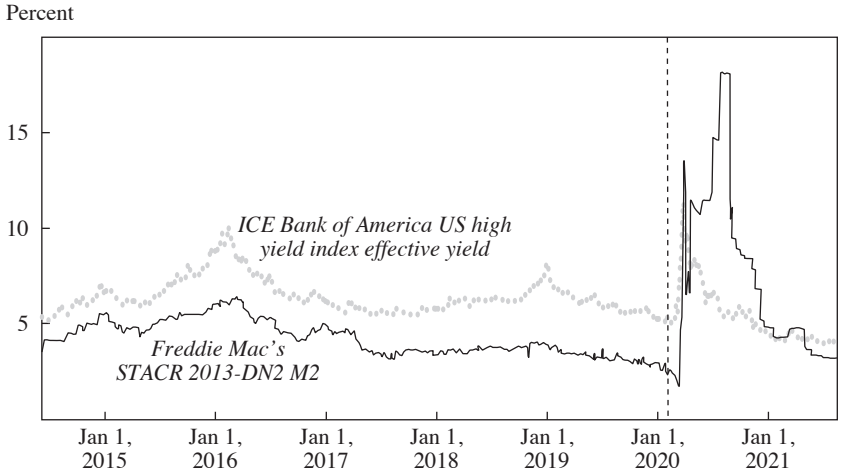
Nonetheless, as figure 3 shows, there was a spike in default fears with the onset of COVID-19.[5] This raises the question of what would have happened if bank and household balance sheets had been as highly levered as they were in the global financial crisis. If there had been an insolvency crisis, say due to high loan-to-value ratios and low equity at the start of the pandemic, together with a pandemic-induced recession, it is arguable whether forbearance would have been sufficient to prevent negative price feedback loops and financial and household debt distress. Regulatory policy

4. Lowered mortgage rates resulted in about $100 billion in lower payments for mortgage borrowers who refinanced (Gerardi and others 2021). See Gerardi, Lowenstein, and Willen (2021) for a discussion of the additional potential benefits of a more streamlined refinance program.

5. The immediate response to COVID-19 shown in trades of credit risk transfers was a large spike in the price of default risk, before the underlying stability of mortgage markets was recognized. The existence of these instruments to price default risk is a post–global financial crisis innovation, as described in Gete, Tsouderou, and Wachter (2022).

15985-03b-Cherry_Comm&GD-4thPgs.indd 214 5/12/22 12:09 PM

Figure 3. Credit Risk Transfers versus High-Yield Bonds



Source: Gete, Tsouderou, and Wachter (2022).

Note: The figure plots the daily spread (yield to maturity minus one month US dollar LIBOR) in the secondary market of the mezzanine tranche M2 of Freddie Mac's STACR CRT security 2013-DN2, and the effective yield index of US high-yield corporate bonds from Bank of America. The vertical line indicates February 1, 2020, which was the onset of COVID-19 in the United States.

and heightened risk concerns would have required lending institutions to halt or slow their lending due to a lack of capital, as they did in the global financial crisis. Borrowers also would have had to deleverage as they did in the global financial crisis. However, due in part to enforced lending standards most homeowners had substantial equity in their homes prior to the pandemic. Dodd Frank policies, including stress tests, provided cushions to bank balance sheets. The substantial government role overseeing the banking system and the GSEs had established, in the years prior to the pandemic, what amounted to an ad hoc macro prudential framework (Hanson, Kashyap, and Stein 2011; Wachter forthcoming).

Arguably, a second outcome of a pervasive public role in mortgage lending was the speed at which the forbearance rules were put into place across mortgage lenders and the rapidity in the delivery of relief. The authors, however, have a different and opposite interpretation of the implementation of forbearance. The authors conclude that the private sector would have granted relief on its own and in substantially the same amount, without government mandates to do so.

The authors show in a difference-in-differences analysis that, at the size break point at which mortgage loans receive government backing,

there is little difference between whether forbearance is granted or not. That is, there is about a 20 percent to one-third higher probability that government-backed loans will be forborne relative to non-government-backed mortgage loans. The authors conclude that most of the forbearance (two-thirds to 80 percent) would have occurred without public action or subsidy that comes from the public sector.

The data certainly demonstrates that loans that did not qualify for government support were also generally forborne. However, the CARES Act requirement, and the Federal Housing Administration (FHA) and GSE regulations put into place in support of the act, likely set uniform standards for and spurred a collective action response that otherwise might not have occurred at all or so quickly. Once these policies for publicly backed mortgage debt were in place, what would have been the reputational implications for large banks, for example, if they had not gone along in their jumbo mortgage lending? For many large and small lenders, would not a standard set of forbearance protocols make sense?

Similarly, a surprising finding that the authors uncover is differences in how nonbank servicers handled forbearance requests even though these were to be automatically granted. Might capacity (size of the entities) help to explain this? Previous research warned that liquidity pressures on thinly capitalized nonbank servicers might well cause a crisis if delinquency overwhelmed the capacity of these servicers.[6] This was a real concern prior to GSE action.[7]

These comments are not meant to question the importance of the paper's findings for the role of forbearance in alleviating illiquidity. The analysis solidly shows that delinquencies declined and by an amount associated with the amount of the forbearance. The authors show that debt forbearance put money into the hands of borrowers when they needed it. Importantly, households who were losing jobs and did not know how they were going

6. Kim and others (2018) point to the limited capital of many servicers and the strains they would be under in the event of a spike in FHA defaults. As argued in Wachter (2018), the fact that such mortgages are backed by the government contains risk; nonetheless, severe capacity and capital constraints could lead to market disruptions. See An and others (2022). Student debt is likely to require further intervention as loan repayment requirements are contemplated.

7. While there was initial uncertainty about such risks and GSE mortgage-backed securities rates spiked, the Federal Reserve stepped in to purchase mortgage-backed securities; in addition, the GSEs' book of business was sufficiently healthy due to the maintenance of lending standards, enabling them to take steps to backstop servicers in this crisis (Golding, Goodman, and Zhu 2021).

15985-03b-Cherry_Comm&GD-4thPgs.indd 216 5/12/22 12:09 PM

to pay their debts were put into a safer position until fiscal and monetary support and the waning of the pandemic could bring the economy back.

REFERENCES FOR THE WACHTER COMMENT

An, Xudong, Larry Cordell, Liang Geng, and Keyoung Lee. 2022. "Inequality in the Time of COVID-19: Evidence from Mortgage Delinquency and Forbearance." Working Paper 21-09. Philadelphia: Federal Reserve Bank of Philadelphia. https://www.philadelphiafed.org/consumer-finance/mortgage-markets/inequality-in-the-time-of-covid-19-evidence-from-mortgage-delinquency-and-forbearance.

Anenberg, Elliot, and Therese Scharlemann. 2021. "The Effect of Mortgage Forbearance on House Prices during COVID-19." FEDS Notes. Washington: Board of Governors of the Federal Reserve System.

Gerardi, Kristopher, Lara Lowenstein, and Paul S. Willen. 2021. "Evaluating the Benefits of a Streamlined Refinance Program." *Housing Policy Debate* 31, no. 1: 51–65.

Gerardi, Kristopher, Laurie S. Goodman, Lauren Lambie-Hanson, Susan M. Wachter, and Paul S. Willen. Forthcoming. "Lessons Learned from Housing Policy during COVID-19." In *Recession Remedies: Lessons Learned from the U.S. Economic Policy Response to COVID-19*, edited by Wendy Edelberg, Louise Sheiner, and Davis Wessel. Washington: Brookings Institution.

Gete, Pedro, Athena Tsouderou, and Susan M. Wachter. 2022. "Pricing Mortgage Stress: Lessons from COVID-19 and the Credit Risk Transfers." Working Paper. Social Science Research Network, March 1. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3613211.

Golding, Edward, Laurie S. Goodman, and Jun Zhu. 2021. "Analysis of the Proposed 2020 FHFA Rule on Enterprise Capital." *Housing Policy Debate* 31, no. 1: 16–32.

Hanson, Samuel G., Anil K. Kashyap, and Jeremy C. Stein. 2011. "A Macroprudential Approach to Financial Regulation." *Journal of Economic Perspectives* 25, no. 1: 3–28.

Kim, You Suk, Steven M. Laufer, Karen Pence, Richard Stanton, and Nancy Wallace. 2018. "Liquidity Crises in the Mortgage Market." *Brookings Papers on Economic Activity*, Spring, 347–428.

Levitin, Adam J., and Susan M. Wachter. 2020. *The Great American Housing Bubble: What Went Wrong and How We Can Protect Ourselves in the Future*. Cambridge, Mass.: Harvard University Press.

Risk Assessment, Data Analysis, and Research Group (RADAR). 2021. *Examining Resolution of Mortgage Forbearances and Delinquencies*. Philadelphia: Federal Reserve Bank of Philadelphia. https://www.philadelphiafed.org/consumer-finance/mortgage-markets/examining-resolution-of-mortgage-forbearances-and-delinquencies-august-2021.

Wachter, Susan M. 2018. Comment on "Liquidity Crises in the Mortgage Market," by You Suk Kim, Steven M. Laufer, Karen Pence, Richard Stanton, and Nancy Wallace. *Brookings Papers on Economic Activity*, Spring, 420–23.

Wachter, Susan M. Forthcoming. "Current Issues for Housing Finance and the Macro Economy." In *Routledge Handbook of Housing Economics*. New York: Routledge.

GENERAL DISCUSSION Robert Hall noted that the COVID-19 recession was different from recessions in the past. The rise in unemployment during the pandemic has been driven by those on temporary layoff, and the economy has since returned to a normal unemployment rate of about 5 percent.[1] While there has been a huge social loss from people being out of work, Hall argued that households did not suffer as much dislocation as they have during past recessions since job losses have not been permanent. According to data in the Current Population Survey, those laid off had a reasonable expectation of returning to their jobs.[2] Indeed, Hall observed that data showed workers did return, which coincided with the implementation of the debt forbearance policy.[3] As such, Hall recommended exercising caution in interpreting the unemployment rate.

Caroline Hoxby considered the effect of forbearance on student loans. In the COVID-19 situation, students who would have gone into default were, instead, automatically put into forbearance or another delayed repayment scheme, she observed.[4] Noting that students do not expect to repay their loans with a high probability in many countries—repayment rates in Brazil and Chile are less than 50 percent, for example—Hoxby wondered whether the intervention in the United States may have created an expectation that student loans will not have to be repaid. The issue is that student loans do

1. Erin Wolcott, Mitchell G. Ochse, Marianna Kudlyak, and Noah A. Kouchekinia, "Temporary Layoffs and Unemployment in the Pandemic," *FRBSF Economic Letter*, Federal Reserve Bank of San Francisco, November 16, 2020, https://www.frbsf.org/economic-research/publications/economic-letter/2020/november/temporary-layoffs-unemployment-pandemic/.

2. Bureau of Labor Statistics, "28.4 Percent of Unemployed Were on Temporary Layoff in December 2020," TED: The Economics Daily, January 13, 2021, https://www.bls.gov/opub/ted/2021/28-point-4-percent-of-unemployed-were-on-temporary-layoff-in-december-2020.htm.

3. Elizabeth Weber Handwerker, Paul B. Meyer, Joseph Piacentini, Michael Schultz, and Leo Sveikauskas, "Employment Recovery in the Wake of the COVID-19 Pandemic," *Monthly Labor Review*, December 2020, Bureau of Labor Statistics, https://www.bls.gov/opub/mlr/2020/article/employment-recovery.htm.

4. US Department of Education, "Department of Education Announces Expansion of COVID-19 Emergency Flexibilities to Additional Federal Student Loans in Default," March 30, 2021, https://www.ed.gov/news/press-releases/department-education-announces-expansion-covid-19-emergency-flexibilities-additional-federal-student-loans-default.

15985-03b-Cherry_Comm&GD-4thPgs.indd 218 5/12/22 12:09 PM

not have any collateral apart from human capital, which is extremely difficult to seize from its possessors.

Hoxby then noted that expectations during the Great Recession were very different from those in the pandemic; during the former many workers who were laid off did not expect to find reemployment at their old jobs.[5] During the pandemic workers underestimated how long they would suffer income losses and how much they would save from reduced consumption or benefit from federal grants, Hoxby suggested.

Addressing Hall and Hoxby, Gregor Matvos remarked that forbearance is not an ideal program in the event of a permanent shock. He wondered, however, whether the pandemic shock was temporary due to the quick government intervention. Debt forbearance during the pandemic may have ensured that the COVID-19 shock, unlike the Great Recession, did not spill into household debt, aggregate demand, or housing prices. According to the paper, the shock was temporary for most people, who saw the forbearance as a credit line. However, 20 percent of the population—those in areas worst hit by COVID-19, those with the worst credit scores, and so on—still hung onto the forbearance, as highlighted by the results in the paper. For such individuals the shock may have been more permanent, Matvos argued. While forbearance would not be effective against a permanent shock in the future, the policy could help prevent temporary setbacks from becoming permanent, he concluded.

Tomasz Piskorski added to Matvos's response, highlighting that a forbearance policy acts as a temporary loan and provides less incentive for individuals to exploit the system, especially when shocks are expected to be transitory. The results in the paper showing that less than 10 percent of eligible mortgage borrowers applied for assistance suggest that the self-selection of the program worked, Piskorski argued. Echoing Pascal Noel's and Susan Wachter's comments, Piskorski agreed that a similar policy may not be effective when the shock is more permanent, though quick implementation of debt forbearance by limiting household debt distress channel may prevent some adverse shocks from having more persistent effects on the economy.

Frederic Mishkin concurred with Piskorski that moral hazard is a significant issue with forbearance programs. The Great Recession and the COVID-19 pandemic were very different crises, Mishkin noted. He

5. Henry S. Farber, "Job Loss in the Great Recession: Historical Perspective from the Displaced Workers Survey, 1984–2010," working paper 17040 (Cambridge, Mass.: National Bureau of Economic Research, 2011), https://www.nber.org/papers/w17040.

15985-03b-Cherry_Comm&GD-4thPgs.indd 219 5/12/22 12:09 PM

observed that during the COVID-19 crisis, the issues people faced repaying their loans did not stem from overborrowing but were a result of the pandemic, which explained the nature of the policies implemented by the Federal Reserve and Congress.[6] Mishkin asked the authors to comment on the lessons learned from the pandemic since it involved very low moral hazard costs. He suggested that a future crisis resulting from overborrowing, for example, might have different policies that would be appropriate.

Donald Kohn reflected on the shadow banking system and nonbank services and their role in a future shock that may be of a less temporary nature than the COVID-19 crisis and would be unlikely to be accompanied by the type of fiscal policy that has provided extraordinary income support during the current recession. Kohn highlighted the need to make the shadow banking system and nonbank services more resilient against future shocks. He then asked the authors to reflect on the importance of differentiating between rent forbearance and mortgage forbearance, mainly due to differences in the nature of the people taking the loans.

Robert Tetlow commented that foreclosure decisions are influenced not just by the duration of the shock but also by its breadth. For instance, given that the COVID-19 shock affected everyone, mortgage lenders had to consider whether there were any alternative borrowers who would be in a better position to pay, leaving less scope to opportunistically capture some of the equity of homeowners. This was particularly relevant for rent forbearance decisions, he noted, because landlords had to consider whether new tenants would be in any better position to pay the rent than their existing tenants.

Steven Davis remarked that, while the temporary layoff share of the unemployed surged during the COVID-19 crisis, the labor force also shrank by roughly 8 million people.[7] The labor force was still below prepandemic levels in August 2021, Davis observed, which implied that many people had not returned to their prepandemic jobs.[8] At the same time, many

6. The Federal Reserve slashed interest rates and greatly expanded lending facilities, while Congress passed legislation to make large payments to households and businesses.

7. Wolcott and others, "Temporary Layoffs"; Nathan Jefferson, "Post-Pandemic Activity Rebounds, but Many Remain Outside the Labor Force," Economic Synopses, Federal Reserve Bank of St. Louis, July 14, 2021, https://research.stlouisfed.org/publications/economic-synopses/2021/07/14/post-pandemic-activity-rebounds-but-many-remain-outside-the-labor-force.

8. US Bureau of Labor Statistics, "Employment Situation News Release," September 3, 2021, https://www.bls.gov/news.release/archives/empsit_09032021.htm.

15985-03b-Cherry_Comm&GD-4thPgs.indd 220 5/12/22 12:09 PM

workers had moved out of the travel and hospitality sector and pursued other employment options, demonstrating that there was a surge in permanent job loss triggered by the pandemic, he concluded.[9]

Piskorski agreed with Mishkin that the nature of the shock is important, suggesting that concern about moral hazard may have been one of the reasons it took almost two years for Congress to implement debt relief during the Great Recession. COVID-19, in contrast, was a canonical example of an exogenous shock, Piskorski observed, which affected a large part of the population and made such concerns less warranted. He argued that there are still important lessons to be learned about the implementation, the speed, the simplicity, and the intermediary frictions related to the COVID-19 forbearance policy as they can inform policy responses to future crises.

Matvos reported that, according to the data collected in the paper, shadow bank services required high amounts of liquidity to pay for forbearance before government-sponsored enterprises stepped in, noting that Kohn's concerns were fair and interesting.

9. Kathryn Dill, "Four Million Hotel, Restaurant Workers Have Lost Jobs. Here's How They're Reinventing Themselves," *Wall Street Journal*, February 22, 2021, https://www.wsj.com/articles/hospitality-workers-are-pushed-into-new-careers-as-pandemic-begins-second-year-11613999342.

KEVIN RENNERT
Resources for the Future

BRIAN C. PREST
Resources for the Future

WILLIAM A. PIZER
Resources for the Future

RICHARD G. NEWELL
Resources for the Future

DAVID ANTHOFF
University of California, Berkeley

CORA KINGDON
University of California, Berkeley

LISA RENNELS
University of California, Berkeley

ROGER COOKE
Resources for the Future

ADRIAN E. RAFTERY
University of Washington

HANA ŠEVČÍKOVÁ
University of Washington

FRANK ERRICKSON
Princeton University

The Social Cost of Carbon: Advances in Long-Term Probabilistic Projections of Population, GDP, Emissions, and Discount Rates

ABSTRACT The social cost of carbon (SCC) is a crucial metric for informing climate policy, most notably for guiding climate regulations issued by the US government. Characterization of uncertainty and transparency of assumptions are critical for supporting such an influential metric. Challenges inherent to SCC estimation push the boundaries of typical analytical techniques and require augmented approaches to assess uncertainty, raising important considerations for discounting. This paper addresses the challenges of projecting very long-term

Conflict of Interest Disclosure: This paper draws on several years of work by the Social Cost of Carbon (SCC) Initiative at Resources for the Future (RFF), an initiative funded by private individual and foundation donors. The work of Adrian E. Raftery and Hana Ševčíková was supported by NIH/NICHD grant R01 HD-070936. Other than the aforementioned, the authors did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this article. They are currently not an officer, director, or board member of any organization with an interest in this article.

15985-04a_Rennert-4thPgs.indd 223 5/12/22 12:08 PM

economic growth, population, and greenhouse gas emissions, as well as calibration of discounting parameters for consistency with those projections. Our work improves on alternative approaches, such as nonprobabilistic scenarios and constant discounting, that have been used by the government but do not fully characterize the uncertainty distribution of fully probabilistic model input data or corresponding SCC estimate outputs. Incorporating the full range of economic uncertainty in the social cost of carbon underscores the importance of adopting a stochastic discounting approach to account for uncertainty in an integrated manner.

As the primary economic measure of the benefits of mitigating climate change, the social cost of carbon (SCC) has been called "the most important number you've never heard of" (*Economist* 2017; Roston 2021). Put simply, the SCC is an estimate, in dollars, of the economic cost (i.e., damages) resulting from emitting one additional ton of carbon dioxide (CO_2) into the atmosphere. Conversely, it represents the benefit to society of reducing CO_2 emissions by one ton—a number that can then be compared with the mitigation costs of reducing emissions. There are analogous metrics for methane (CH_4) and nitrous oxide (N_2O). The SCC has deep roots in economics. Indeed, many textbooks use carbon emissions and the resulting climate change as the canonical example of an externality that must be addressed through Pigouvian taxation or other means to maximize human welfare. In particular, basic economic theory recommends that an optimal tax on CO_2 emissions (a carbon tax) be set equal to the SCC, for which marginal damages are measured along an optimal emissions trajectory (Pigou 1920; Nordhaus 1982).[1]

But the relevance and application of the SCC go well beyond its role in determining an optimal Pigouvian tax. As political leaders and stakeholders debate both the broad outlines and the fine details of policies to reduce carbon dioxide emissions, the SCC lies in the background as a remarkably important calculation, used by the US federal government for more than a decade for developing vehicle fuel economy standards and power plant emissions rules. Such analyses have been a mainstay of the regulatory rule-making process since Executive Order 12291 was issued more than forty years ago.[2]

1. This result derives from a simple model lacking many real-world complications such as leakage, tax interaction effects, and other market distortions like research and development (R&D) spillovers, but it represents a reasonable approximation.

2. Executive Order 12291 was the original Reagan-era guidance for benefit-cost analysis, later superseded by Executive Order 12866 in 1993.

15985-04a_Rennert-4thPgs.indd 224 5/12/22 12:08 PM

The SCC has also been the basis for the value of federal tax credits for carbon capture technologies, beginning in 2018 (Rodgers and Dubov 2021), and zero-emissions credits for nuclear power in New York State.[3] The power grid operator for New York is working to include the SCC as a cost adder on top of energy supply bids submitted by power plants, thereby reflecting social costs into market prices and plant dispatch.[4] Many other states have used the SCC as the basis for climate policies and as a benchmark against which proposed carbon prices are compared.[5] Proposed applications include federal procurement decisions and royalties on oil and gas leases on federal land (Prest 2021; Prest and Stock 2021; White House 2021b, sec. 5[b][ii]).[6]

Construction of the SCC and the benefits of reducing emissions are also somewhat distinct from the distribution of benefits. That is, because the consequences of climate change will be different for different communities (country, region, income, social identity), the benefits of mitigating climate change will similarly vary. For example, rising temperatures are likely to create heavier burdens on already hot (and often poor) countries like Bangladesh than on cold (and often rich) countries like Norway. Putting greater weight on dollar-value effects in poorer communities—that is, equity weighting (Errickson and others 2021)—is not the current standard practice, however. Rather, the distribution of effects (when available) is presented alongside the aggregate, unweighted summary. Weighting becomes important as we gain understanding of the distribution of effects.

Estimation of the SCC goes back to William Nordhaus (1982) and has recently seen increasing prominence. In 2018, the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel was awarded to Nordhaus (alongside Paul Romer) for his work incorporating climate

3. State of New York Public Service Commission, Case 15-E-0302 and Case 16-E-0270, "Order Adopting a Clean Energy Standard," https://documents.dps.ny.gov/search/Home/ViewDoc/Find?id=<44C5D5B8-14C3-4F32-8399-F5487D6D8FE8>&ext=pdf, page 131.

4. New York ISO, "Carbon Pricing," https://www.nyiso.com/carbonpricing.

5. Institute for Policy Integrity, The Cost of Carbon Pollution, "States Using the SCC," https://costofcarbon.org/states; Resources for the Future, "Carbon Pricing 101," https://www.rff.org/publications/data-tools/carbon-pricing-bill-tracker/; see also Johnson (2009).

6. Many aspects of climate policy decisions are not necessarily tied to the SCC. Essentially, these include all policy design issues *beyond* measuring benefits and balancing with costs, such as optimal R&D spending amid knowledge spillovers, cost-effective policy design (e.g., uniform standards versus flexible incentive-based policies), interactions between policies (Goulder 1995; Barrage 2020a, 2020b; Borenstein and others 2019), and differences in the distribution of the costs (and in certain cases government revenues) associated with different policy approaches. These are distinct from the question of estimating the marginal benefits of reducing emissions.

15985-04a_Rennert-4thPgs.indd 225 5/12/22 12:08 PM

change into economic analysis, including the role of the SCC in informing policy.[7]

The SCC is typically estimated using integrated assessment models, such as the Dynamic Integrated Climate Change (DICE) model developed by Nordhaus. Integrated assessment models couple climate and economic models to estimate the economic effect of an incremental pulse of CO_2 emissions (in tons) on climate and economic outcomes. The net present value of changes in economic outcomes, divided by the number of tons in the pulse, delivers the SCC. However, many integrated assessment models used in SCC estimates have not kept up with rapidly evolving climate, economic, and demographic science. Moreover, as Nordhaus (1982) noted, many of the factors underlying the SCC are deeply uncertain—notably, our understanding of Earth's climate, the effect of climate change on economic outcomes, and future socioeconomic conditions that capture the discounted consequences from changes in emissions today. The need for robust policy decisions implies we should update the SCC over time to refine central estimates and the range of uncertainty as our scientific understanding progresses.

In this paper, we review efforts to update determinants of the SCC to reflect the best available science, based on the recommendations of a 2017 committee report by the National Academies of Sciences, Engineering, and Medicine (NASEM 2017). This updating is particularly relevant in light of Executive Order 13990 (January 20, 2021), which reestablished the Obama-era Interagency Working Group (IWG) on the Social Cost of Greenhouse Gases and directed it to update the SCC. We also note other research efforts on updating the SCC.

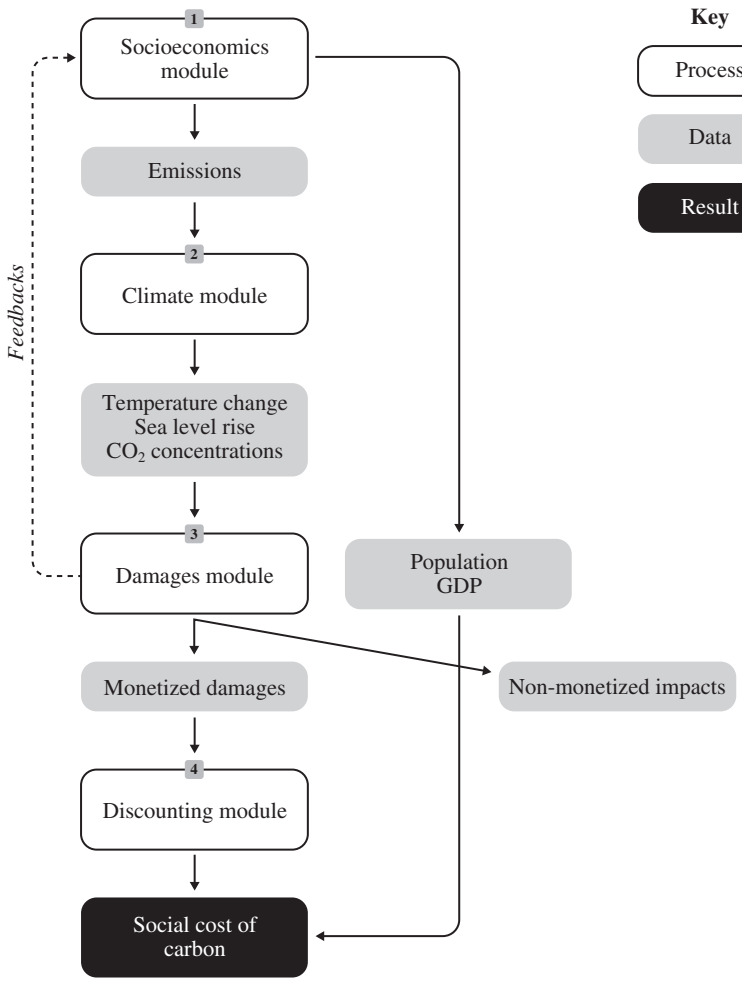
The NASEM report recommended creating an integrated framework comprising four components ("modules") underlying the SCC calculation: *socioeconomics*—probabilistic projections of population, gross domestic product (GDP), and emissions over multiple centuries; *climate*—an improved model of Earth's climate system and climate change; *damages*—the economic consequences of climate change, based on recent studies; and *discounting*—aggregated present-value marginal damages and stochastic discount factors that correctly reflect the uncertain socioeconomic drivers.

Figure 1 shows how the modules fit together, including how socioeconomics affects emissions trajectories, which are input into the climate

7. The Nobel Prize, "The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2018," https://www.nobelprize.org/prizes/economic-sciences/2018/summary/.

15985-04a_Rennert-4thPgs.indd 226 5/12/22 12:08 PM

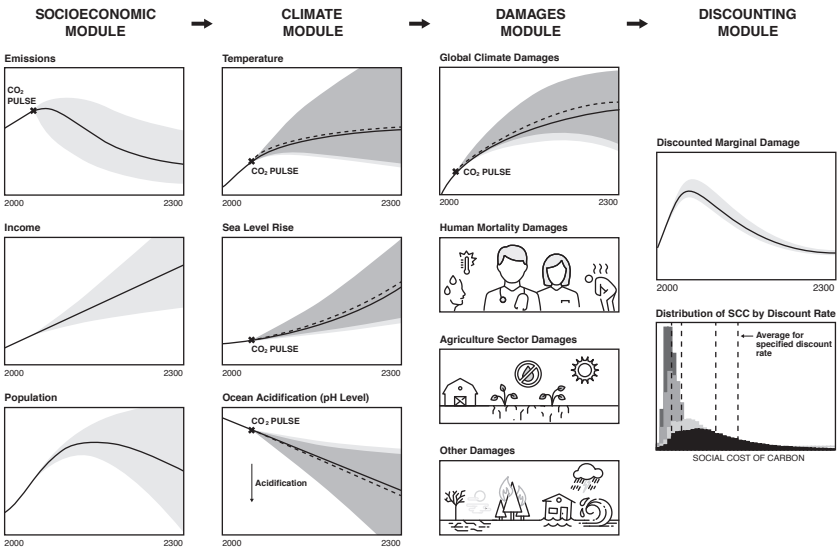
Figure 1. Modularized Approach to Estimating SCC



Source: National Academies of Sciences, Engineering, and Medicine (2017). Adapted and reproduced with permission from the National Academy of Sciences, courtesy of the National Academies Press, Washington, DC.

Note: The damages module may require regional and/or sectoral socioeconomic and climate data either as direct inputs or for calibration.

Figure 2. Estimating Social Cost of Carbon under Uncertainty



Source: Authors' calculations.

Note: Estimation involves a baseline case (solid and shaded area) versus a pulse of emissions (dashed lines and shaded area). Shading depicts probability distributions on projections.

module to project future temperatures. These temperatures are converted into a stream of future economic losses in the damages module (also influenced by socioeconomic trajectories), which are then discounted to a present value in the discounting module.

Because the SCC represents the marginal effect of an incremental ton of emissions, this entire model is run twice—once as a baseline and once with a small pulse of additional emissions (figure 2). The resulting change in the stream of economic damages per ton from this emissions pulse, in present value, is the SCC. More generally, when inputs to a module are uncertain (e.g., because of uncertainty about the climate's response to emissions or about future economic growth), modelers have incorporated that uncertainty through Monte Carlo analyses by taking draws of (potentially correlated) probability distributions of each random variable. The result is a distribution of SCCs, often summarized by its expected value. For example, the federal government's current interim value of \$51/ton CO_2 (IWG 2021) reflects the expected value of the SCC over uncertainty in the climate's warming response and scenarios of economic growth and population, at a 3 percent constant discount rate.

The NASEM report noted that the IWG's estimates of the SCC, including the current interim $51/ton SCC value, used somewhat dated and often simplistic modules. For example, five socioeconomic scenarios were not developed with formal probabilities attached but were treated as equally likely. The scenarios did not incorporate the work done by economists, demographers, and statisticians to estimate and quantify uncertainty around long-term economic and population growth. Also, the discounting approach used a constant discount rate rather than treating the discount rate as stochastic; that choice becomes increasingly important as the decision horizon extends into the future. The IWG noted the potential for a declining term structure and correlation between the discount rate and damage outcomes but did not consider an explicit stochastic discount factor that accounts for both future discount rate uncertainty and, through uncertain socioeconomic outcomes, correlation with the damages being discounted. To address such shortcomings, the NASEM report issued recommendations for improvement, which Executive Order 13990 specifically directed the IWG to consider.

This paper documents recent work that has improved the scientific basis for the modules so that the IWG can update the SCC to reflect the best available science. Section I discusses the improved socioeconomic module, with long-term probabilistic projections of population, economic growth, and emissions. Section II illustrates how an incremental ton of emissions translates into climate and economic effects (damages). Section III discusses the crucial role of the discount rate, given recent research on declining equilibrium interest rates, plus the importance of using stochastic discount factors and the shadow price of capital for valuing effects on investment. Section IV then combines these elements into a simplified model of the SCC, with associated uncertainty bounds for the socioeconomic, climate, damages, and discounting components. Finally, section V concludes and raises issues that await future research.

I. Economic and Demographic Drivers of Climate Effects

Assessments of damages from climate change are influenced by projections of population, economic growth, and emissions. Population growth can drive emissions and increase or decrease total economic exposure to the health effects of climate change. Economic growth similarly affects both the level of expected emissions and the resulting damages, which are often estimated to scale with economic activity (Diaz and Moore 2017). For example, the monetization of mortality consequences typically depends on per capita income (Robinson, Hammitt, and O'Keeffe 2019). Economic

growth projections can also influence the SCC through the discount rate if estimates are calculated using Ramsey-like discounting, where the discount rate is a function of the rate of economic growth: higher (lower) growth scenarios will yield a higher (lower) discount rate. Finally, projections of global emissions determine the background state of the climate system against which damages from an additional pulse of emissions are measured.

Estimates of the SCC are highly sensitive to socioeconomic and physical projections (Rose, Diaz, and Blanford 2017), but revised estimates have been based primarily on changes in socioeconomic projections, not on improved understanding of the climate system (Nordhaus 2017b). Explicitly considering realistic, probabilistic socioeconomic projections is thus important for improving the characterization of both the central tendency and the uncertainty in the SCC.

A robust characterization of socioeconomic contributions to SCC estimates would ideally incorporate probabilistic projections of population, economic growth, and emissions. The particular requirements of SCC estimation, however, pose significant challenges for generating such projections. One is the time horizon: given the long-lived nature of greenhouse gases in the atmosphere, the SCC needs to account for discounted damages two hundred to three hundred years into the future (NASEM 2017). Yet nearly all projections, such as the scenarios previously used by the IWG (2010) and the shared socioeconomic pathways used by the IPCC (Riahi and others 2017), end at year 2100 and are often scenario-based rather than probabilistic. New probabilistic projections that extend well into the future are required.

Another challenge is that although climate change can be projected from emissions scenarios consistent with globally aggregated projections of economic activity and population growth, the resulting climate damages are most appropriately estimated at a regional (or even local) scale. Thus, they require geographically disaggregated estimates of GDP and population.

A third challenge is that the future path of emissions likely depends on uncertain improvements in technology and on the scale and success of policy interventions outside the range of the historical record. That is, whereas historical data may be a reasonable guide to forecast population and economic activity, the same is not true for emissions. The SCC should be measured against our best estimate of future emissions, inclusive of future mitigation policies *except* the one under analysis.

The fourth issue is the interrelated nature of these variables: the projections for each variable must be consistent with one another. For example,

emissions intensity might be lower with higher economic growth (and its associated wealth and technological improvements).

I.A. Past Approaches to Socioeconomic Projections

In lieu of using fully probabilistic socioeconomic projections, researchers have typically turned to socioeconomic scenarios, which can provide consistency across analyses and still incorporate specific narratives. The IWG adopted a scenario approach in its initial estimates (IWG 2016), and these same scenarios support the interim estimates put forward by the Biden administration in January 2021 (IWG 2021). The five socioeconomic scenarios were drawn from the Energy Modeling Forum 22 exercise (Clarke and Weyant 2009), selected to span roughly the range of emissions outcomes in the full set of the forum's scenarios and thus represent uncertainty across potential socioeconomic projections. Only one of the scenarios represented future climate policy. The IWG extended the five scenarios to 2300 by assuming that GDP and population growth each decreased linearly to zero in 2300. The five scenarios were assigned equal probability for computing an expected value for the SCC (no such probabilistic interpretation existed for the work by the Energy Modeling Forum 22).

The IWG scenarios were critiqued for not spanning the uncertainty in a full set of relevant socioeconomic variables (e.g., GDP, population) or reflecting the broader scenario literature overall (Rose and others 2014; Kopp and Mignone 2012). The resulting SCC estimates, then, may not reflect damage calculations based on the full range of expected variation. The NASEM panel noted that the IWG did not provide a rationale for its scenario weighting or the choice to extend the scenarios from 2100 to 2300 by assuming that GDP and population growth each decreased linearly to zero. The panel recommended using a combination of statistical methods and expert elicitation to generate a set of probabilistic long-term projections for each variable.

Subsequently, a multidisciplinary research effort developed the shared socioeconomic pathways (SSPs) (Riahi and others 2017), scenarios intended primarily to support the assessment efforts of the Intergovernmental Panel on Climate Change (IPCC). Each of the five SSPs consists of quantified measures of development and an associated narrative describing plausible future conditions that drive the quantitative elements. The SSPs end in 2100, but researchers have offered extensions to 2300 (Nicholls and others 2020; Kikstra and others 2021). The SSPs are freely available and comprehensive, have an extensive publication record, and are expected to be used in

the IPCC's *Sixth Assessment Report*. For these reasons, we use the SSPs as our primary point of comparison.

Scenarios in general, and the SSPs in particular, do not come (as the IWG assumed) with associated probabilities. That limits their utility in evaluating uncertainty. Although the SSP authors have themselves cautioned against using the SSPs in a probabilistic fashion, Ho and others (2019) sought to address this limitation through an expert survey assessing the likelihood of each SSP. Others have sought to guide scenario usage by characterizing the plausibility of various scenarios (Stammer and others 2021). Even without formal probabilities, in practice, the SSPs are often interpreted in modeling exercises as representing the uncertainty between high-emissions (SSP5) and low-emissions (SSP1) futures, at times with the implication that the difference represents a "no policy" counterfactual versus a "likely policy" scenario. This has led to a recent debate over the viability of the high-emissions scenario, given the current pace of technology evolution, among other factors (Hausfather and Peters 2020).

Previous efforts to quantify the uncertainty of socioeconomic projections over a century are limited. Raftery and others (2017) used a statistical approach to generate density functions of country-level economic growth per capita, population, and carbon intensity (CO_2/GDP) to project a density of future emissions trajectories via the IPAT equation (Commoner 1972), similar to our socioeconomic approach.[8] Müller, Stock, and Watson (2020) employed a Bayesian latent factor model that projects long-run economic growth based on low-frequency variation in the historical data of country-level GDP per capita.[9] Christensen, Gillingham, and Nordhaus (2018) conducted an expert survey of economists to quantify the 10th, 50th, and 90th percentile ranges of economic growth for six groupings of countries. Comparing results with the SSP ranges, they found that the SSPs underestimated the range of uncertainty expected by the experts and that using the increased range for economic growth with the DICE model suggested that emissions were also underrepresented by the SSPs.

The NASEM (2017) report noted that statistical models based solely on historical data are unlikely to fully inform the variability of future projections over centuries, suggesting caution in using raw outputs from statistical

8. The IPAT equation is Impact = Population × Affluence × Technology, a heuristic for thinking about the impact of humans on the environment.

9. The method used by Müller, Stock, and Watson (2020) extends the approach provided in Müller and Watson (2016), which was suitable only for global estimates of economic growth, to generate internally consistent growth projections at the country level.

15985-04a_Rennert-4thPgs.indd 232 5/12/22 12:08 PM

models over long time scales. This concern led the NASEM panel to recommend using formal expert elicitation to quantify the uncertainty around future long-run projections, which can then be used to augment projections from statistical models.

We next describe efforts undertaken by the Resources for the Future's (RFF) Social Cost of Carbon Initiative and collaborators to build on both statistical and expert-based approaches to generate distributions of projections of population and GDP per capita at the country level, plus distributions of the three primary greenhouse gases (CO_2, CH_4, and N_2O) at the global level. The resulting probabilistic distributions, collectively referred to as the RFF Socioeconomic Projections (RFF-SPs), fully incorporate the NASEM recommendations for generating an improved socioeconomic module for SCC estimation.

I.B. Probabilistic Population Projections to 2300

METHODS To develop probabilistic, country-level population projections through 2300, we start with the fully probabilistic statistical approach that has been used since 2015 by the United Nations (UN) for its official population forecasts to 2100 (United Nations 2019). We then extend the statistical model to 2300, incorporating feedback and improvements suggested by a panel of nine leading demographic experts that we convened to review preliminary results. This work is detailed in Raftery and Ševčíková (2021).

The UN uses a probabilistic method built on the standard deterministic cohort-component method of population forecasting (Preston, Heuveline, and Guillot 2001). This method projects forward the three components of population change: fertility, mortality, and migration, broken down by age and sex. The probabilistic method builds Bayesian hierarchical models for each of the three components and projects them forward probabilistically using a Markov chain Monte Carlo method, which produces a large number of trajectories (typically 1,000–2,000) of future numbers of births, deaths, and migration events in each country by age and sex. Each trajectory of fertility, mortality, and migration is then combined to give a trajectory of future population by age and sex in each country. These trajectories of population numbers in turn approximate a probability distribution for any population quantity of interest (Raftery and others 2012; Raftery, Alkema, and Gerland 2014; Gerland and others 2014).

Fertility is projected by focusing on each country's total fertility rate (TFR), which is the expected number of children a woman would have in a given period if she survived the reproductive period (typically to age 50)

and at each age experienced the age-specific fertility rates of that period. The UN models the evolution of fertility in all countries using a Bayesian hierarchical model that divides it into three phases depending on where it lies in the fertility transition from high to low fertility (pre-transition, transition, post-transition). It then fits a time series model to each phase, accounting for spatial correlation between countries (Alkema and others 2011; Raftery, Alkema, and Gerland 2014; Fosdick and Raftery 2014; United Nations 2019; Liu and Raftery 2020).[10] Mortality is similarly projected by focusing on life expectancy at birth.[11] This is projected by another Bayesian hierarchical model for all countries for both sexes (Raftery and others 2013; Raftery, Lalic, and Gerland 2014). The UN has traditionally projected net international migration for each country deterministically by assuming that it would continue in the future at the same rate as currently (United Nations 2019).

We extended the UN's method, designed for projections to 2100, out to 2300 and preliminary results were reviewed by a panel of nine expert demographers that we convened.[12] While broadly supportive, the panelists were in agreement that the resulting uncertainty bounds for TFR in 2300 were too narrow and that in particular the lower bound of the 95 percent prediction interval for world TFR in 2300 (1.66) was too high. A lower bound of 1.2 children per woman for the world TFR in 2300 was suggested as a more plausible lower bound. We incorporated this recommendation by adding a worldwide random walk component to the TFR model.

Experts on the panel also suggested that international migration should be projected probabilistically, in line with the general approach, rather than deterministically as done by the UN. We implemented this by projecting net international migration using a Bayesian hierarchical model (Azose

10. The TFR has evolved in a similar way in all countries. In preindustrial times, the TFR for a typical country was high (in the range of 4–8 children per woman). Then, usually after the onset of industrialization, it started to decrease. After a bumpy decline lasting several decades to a century, the TFR flattened out at a level below the replacement rate of about 2.1 children per woman. This decline is called the fertility transition. After the end of the fertility transition, the TFR fluctuated without a clear trend, mostly staying below the replacement rate. For example, in the United States, the TFR was around 7 children per woman in 1800 and then declined, reaching 1.74 in 1976 and thereafter fluctuating up and down; it is now 1.64, close to the level it was at in 1976 (Raftery 2021).

11. The general trend since 1840 has been that life expectancy has increased steadily (Oeppen and Vaupel 2002), with slower increases for countries with the lowest and highest life expectancy and the fastest increases for countries in the middle.

12. Each panelist provided written reviews of the preliminary projections and methodology, and all except Tomáš Sobotka presented them as part of a virtual workshop convened by Resources for the Future on October 4, 2018. Panelists are listed in the acknowledgments.

15985-04a_Rennert-4thPgs.indd 234 5/12/22 12:08 PM

and Raftery 2015; Azose, Ševčíková, and Raftery 2016). We additionally implemented the final panel recommendation to impose constraints on population density to prevent unrealistically high or low population numbers in some age groups in some countries.

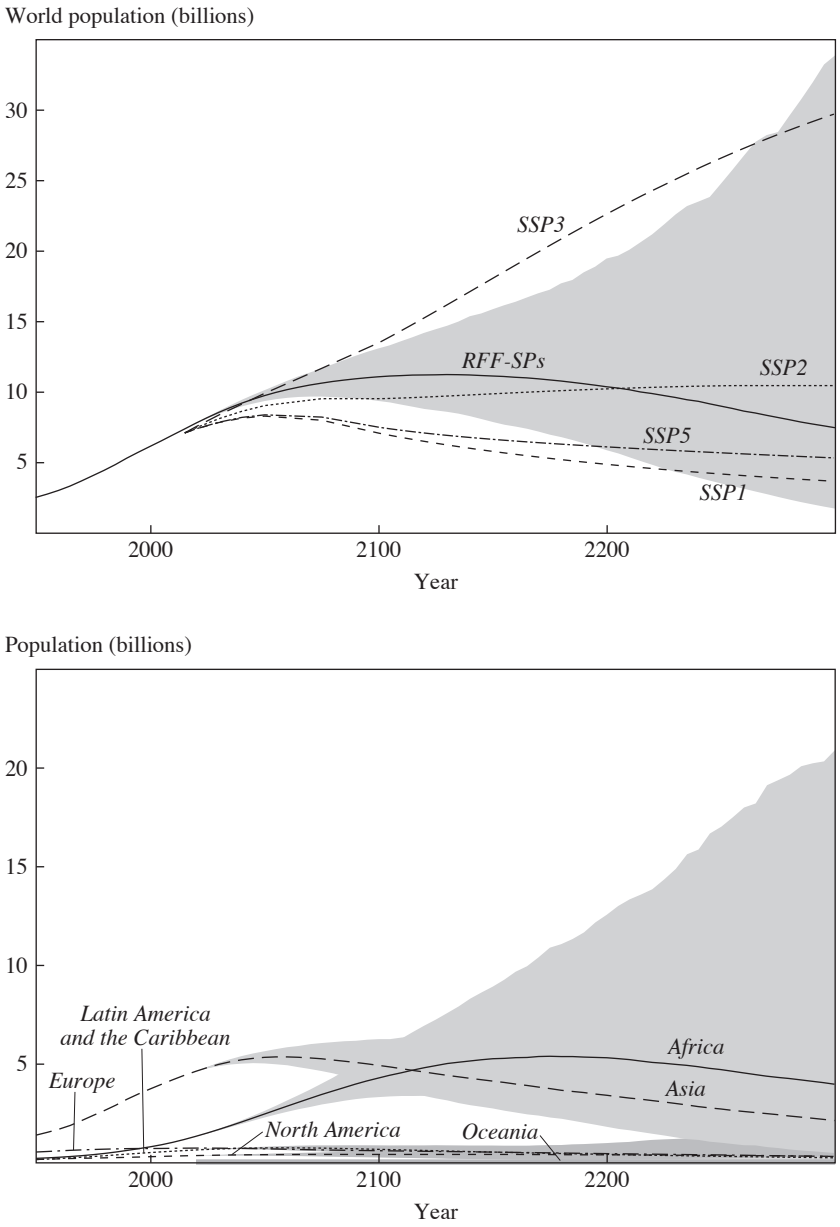
RESULTS The resulting population projections for 2300 for the world as a whole and for the continents are shown in figure 3. They show that total world population is likely to continue to increase for the rest of the twenty-first century, albeit at a decreasing rate, to level off in the twenty-second century, and to decline slightly in the twenty-third century. Uncertainty for 2300 is considerable, appropriately, reflecting the very long forecast time horizon, with a median forecast of 7.5 billion, but a 90 percent interval from 2.8 to 20.5 billion. The results agree closely with the UN forecasts for the period to 2100 (United Nations 2019).

Figure 3 also shows the results for each major continental region. The populations of Asia, Europe, and Latin America are likely to peak well before the end of this century and then decline substantially. The populations of Africa and North America are also likely to peak and then decline but much later, in the twenty-second century. In the case of Africa this is due to population momentum (with a high fraction of the population currently in reproductive ages) and current high fertility. In the case of North America it is due to a combination of modest population momentum, fertility that is closer to replacement level than in other continents, and immigration. Uncertainty for each region in 2300 is high.

In comparison to the population projections from the SSPs, our population projections are centered around a peak of slightly over 10 billion people globally reached late this century, lying closest to SSP2, although SSP2 levels off at a higher level than our median projection after 2200. Through 2300, the 90 percent confidence distribution around our median is narrower than the range indicated by the SSPs and considerably narrower through 2200. SSP1 and SSP5 lie below the 5th percentile of our distribution through almost the entire time horizon to 2300. SSP3 features a very aggressive population projection in the top tail of the distribution, at about the 99th percentile in 2300. In sum, none of the SSPs has a central tendency for population in line with our fully probabilistic projections, and the range of population given by SSP1–SSP5 is wide relative to ours.

We are aware of only three other detailed efforts to project world population to 2300, all of them deterministic, in contrast with our probabilistic method described here. One was carried out by the United Nations (2004) and was deterministic but containing several scenarios. The range of these projections for 2300 from the different scenarios went from 2.3–36.4 billion,

Figure 3. Probabilistic Population Projections for World and Major Regions, to 2300



Source: Authors' calculations based on Raftery and Ševčíková (2021).

Note: Data prior to 2020 are from the UN's *World Population Prospects 2019*. The predictive medians are shown as solid curves; the shaded areas show the 90 percent and 98 percent predictive intervals. The world population projections from the extended SSPs are shown for comparison.

compared with our 98 percent prediction interval of 1.7–33.9 billion. Although using different methodologies and carried out over fifteen years apart, the two sets of projections give results that are compatible with one another, perhaps to a surprising extent.[13]

Another such exercise was carried out by Vallin and Caselli (1997), also deterministic with three scenarios corresponding to different long-term trajectories of world TFR. Two of the scenarios led to world population stabilizing at around 9 billion, while the other resulted in 4.3 billion people in 2300. All three of these scenarios give world population in 2300 well within our 80 percent interval, though with a range that is much narrower than either ours or that of United Nations (2004). Gietel-Basten, Lutz, and Scherbov (2013) also performed a projection exercise to 2300, with a very wide range of scenarios for long-term world TFR. They obtained projections of global population yielding anything from zero to 86 billion in 2300.[14]

I.C. Probabilistic Economic Growth Projections to 2300 and Economic Growth Survey

METHODS The probabilistic projections of economic growth often used in analyses by governments and the private sector have not incorporated the time scale of centuries, as is needed to support SCC estimates and other economic analyses of climate change. Müller, Stock, and Watson (2020) took a significant step forward by providing probabilistic econometric projections over long periods. Their methodology involves a multifactor Bayesian dynamic model in which each country's GDP per capita is based on a global frontier of developed economies (countries in the OECD) and country-specific deviations from that frontier. Correlations between countries are also captured in a hierarchical structure that models countries in "covariance clubs," in which country-level deviations from the frontier vary together. The hierarchical structure also permits pooling information across countries, an approach that tightens prediction intervals. This model is then estimated on data for 113 countries over 118 years (1900 to 2017). The model yields 2,000 sets of trajectories of country-level GDP per capita from 2018 to 2300. Each can be considered an equally likely uncertain future. Each is characterized by a path for the global factor and 113 country-specific deviations from that pathway. The results are described more

13. The very high upper bound for the UN (2004) projections is likely an artifact of the perfect correlation implied by the deterministic scenarios and the aggregation of such results.

14. As in the UN (2004) projections, these very extreme outcomes are likely due in part to the perfect correlation between countries implied by the deterministic scenarios and the aggregation of such results.

fully below; for more information about the model, see Müller, Stock, and Watson (2020).

As noted earlier, however, NASEM (2017) recommended augmenting statistical models with formal expert elicitation to quantify uncertainty, especially for long-term projections. But surveying experts on long-term uncertainty of economic growth at the country level is impractical because of time constraints and the difficulty of accounting for intercountry correlations. Consequently, our study was designed to work in tandem with an econometric model that provides country-level projections and represents the intercountry dynamics. The RFF Economic Growth Survey focused on quantifying uncertainty for a representative frontier of economic growth in the OECD countries. The results informed econometric projections based on the model by Müller, Stock, and Watson (2020) of an evolving frontier (also based on the OECD), in turn providing country-level, long-run probabilistic projections.

The methodology we applied is the "classical model" (Cooke 1991, 2013) of structured expert judgment, analogous to classical hypothesis testing. In essence, the experts are treated as statistical hypotheses: they are scored on their ability to assess uncertainty based on their responses to calibration questions whose true values are known to us but unknown to the experts. This scoring allows us to weight the experts' judgments, and the scores of combinations of experts serve to gauge and validate the combination that is adopted. The ability to performance-weight experts' combined judgments has generally been shown to provide the advantages of narrower overall uncertainty distributions with greater statistical accuracy and improved performance both in and out of sample (Colson and Cooke 2017, 2018; Cooke, Marti, and Mazzuchi 2021).

Ten experts, selected for their expertise in macroeconomics and economic growth and recommended by their peers, were elicited individually by videoconference in roughly two-hour interviews in 2019–2020. They received an honorarium where appropriate. The full elicitation protocol is available in the online appendix; the general process was as follows. First, experts quantified their uncertainty for several initial questions, after which answers were provided for self-assessment; this step was intended to familiarize them with the process and alert them to potential biases. The experts then provided a median and 90 percent confidence range for eleven calibration questions for which the true values were known to us.

Experts next provided their 1st, 5th, 50th, 95th, and 99th quantiles for the variables of interest: levels of OECD GDP per capita for 2050, 2100, 2200, and 2300. For experts more comfortable working with growth rates

15985-04a_Rennert-4thPgs.indd 238 5/12/22 12:08 PM

(rather than levels), we provided a spreadsheet tool that translated average growth rates into GDP per capita levels. The experts were informed that their combined quantiles of GDP levels would be further combined with country-level econometric projections, as described below, but they were not shown the results. They were given historical data on economic growth to provide a consistent baseline of information across the panel, and they were permitted to consult outside sources if desired. The experts provided additional rationale for their quantiles verbally throughout the elicitation and concluded the survey by formally identifying the primary factors driving their low and high future growth scenarios.

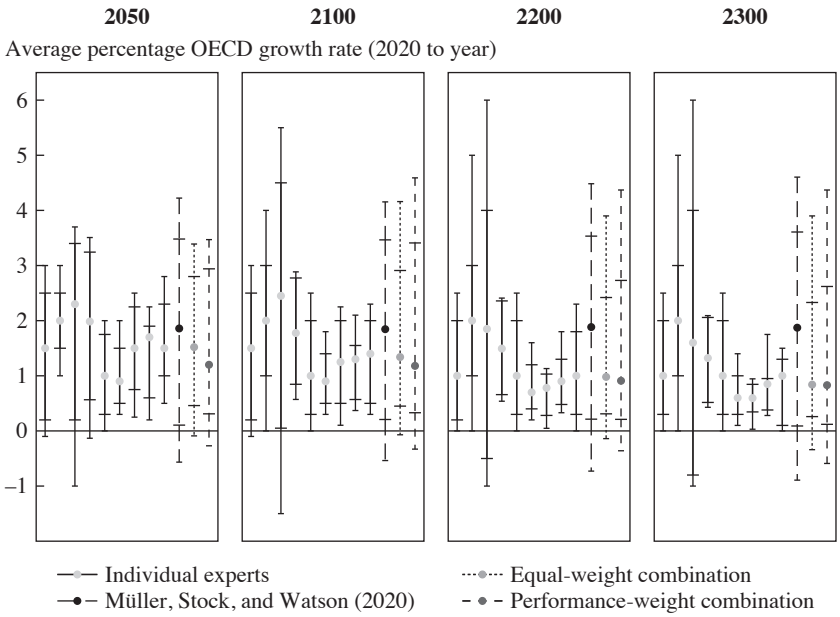
Given that the projections were being used as an input to the estimation of climate change damages, which would reduce economic activity below the projected level, the experts were specifically asked to provide quantiles of economic growth absent the effects of further climate change as well as absent further policy efforts to reduce emissions. Two of the ten experts provided a pair of modified base quantiles to reflect the absence of effects from climate damages and climate policy that are utilized here, but in general the proposed modifications to their original distributions were minor. Moreover, several experts noted that although climate change was a primary factor underlying their probability of low growth projections, the complexity of the multiple uncertain factors represented in their base quantiles precluded systematic removal, and they deemed their base quantiles appropriate for assessing uncertainty in the SCC and other analyses assessing the economic damages from climate change.

The results of the expert elicitations were combined by first fitting each expert's five quantiles for each year, in log GDP per capita, with a Johnson S_U distribution (Johnson 1949) to generate a continuous cumulative distribution function specific to each expert. We next combined the cumulative distribution functions in two ways: averaging across the set of expert functions with equal weight, and performance-weighting the experts according to their performance on the calibration questions. This process yielded a pair of final combined elicited values of OECD GDP per capita for each elicited year and quantile.[15]

RESULTS OF ECONOMIC GROWTH SURVEY On the calibration questions (see online appendix), the experts demonstrated an overall high level of statistical accuracy compared with other structured expert judgment studies and results that are robust against expert loss. As shown by their individual quantiles (figure 4) and as expressed in comments during the videoconferences, most

15. See online appendix for further detail.

Figure 4. Distributions of Future Average GDP Growth Rates for OECD



Sources: RFF Expert Growth Survey; Müller, Stock, and Watson (2020); and authors' calculations.
Note: For each bar, the circle shows the median and the lines show the 1st, 5th, 95th, and 99th percentiles of the relevant distribution.

participants' median forecast was that long-term growth would be lower than the growth rate of the past one hundred years. The responses show considerable diversity in their characterization of uncertainty around the median, however, with some of the widest ranges being driven by their explicit inclusion of events that are not present or fully realized in the historical record of economic growth on which statistical growth projections are based.[16] When asked to identify the primary drivers of the low-growth quantiles, the experts most commonly responded with climate change, followed by world conflict, natural catastrophes, and global health crises. Rapid advancement of technology was cited most often as the primary driver of high growth, followed by regional cooperation and advances in medical science. Many experts expected that technology breakthroughs in

16. The quantiles from one expert included global civilization-ending events that were outside the scope of the survey and incompatible with assumptions for US federal policy analysis; they unreasonably distorted the combined distributions toward extreme values. Quantiles from this expert were excluded in the final survey.

15985-04a_Rennert-4thPgs.indd 240 5/12/22 12:08 PM

clean energy would dramatically lower global emissions. Implicit in this narrative is a negative correlation between economic growth and carbon dioxide emissions.

As shown in figure 4, both the performance-weighted and the equal-weighted combinations of the experts' distributions yield narrower ranges as well as lower medians than do the statistical trajectories for all four years (2050, 2100, 2200, and 2300). The median of the equal-weighted combination is consistently higher than the median based on performance weighting, but the difference shrinks throughout the period until the medians nearly converge in 2300. Overall, the experts viewed sustained long-term growth rates above 4 percent or even slightly below zero percent as highly unlikely but not impossible.

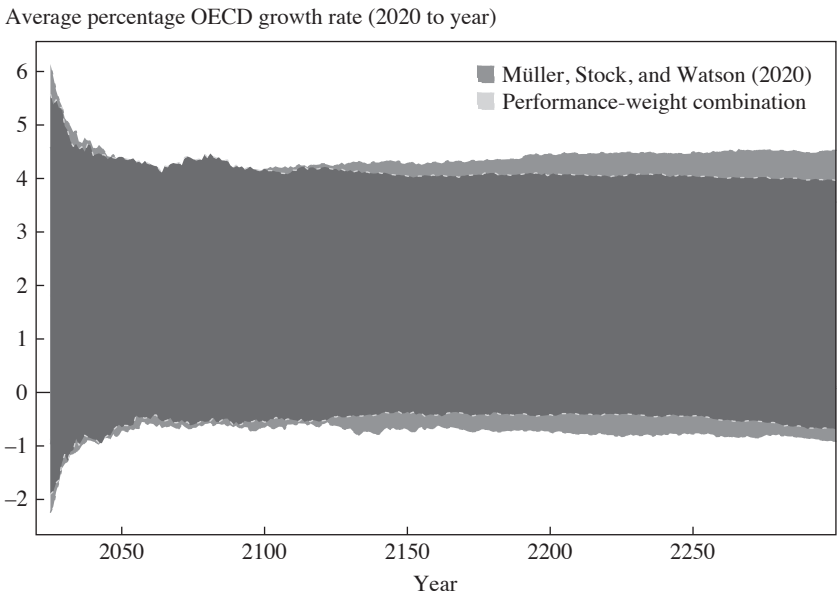
RESULTS OF ECONOMETRIC GROWTH PROJECTIONS AUGMENTED WITH EXPERT JUDGMENT We used the survey results to modify econometric projections of GDP per capita based on the methodology of Müller, Stock, and Watson (2020) and to generate density functions of internally consistent projections of economic growth at the country level. As indicated in Müller, Stock, and Watson (2020), economic growth 100–300 years into the future is highly uncertain, well beyond that captured in typical scenario projections (see figure 5).

The tails of the Müller, Stock, and Watson (2020) distribution are quite wide, leading to some implausibly small or implausibly high long-term average growth rates in the extreme tails (e.g., below the 1st percentile or above the 99th percentile). These extreme tails correspond to extremes of persistent economic growth beyond what has been observed historically over long periods (e.g., below –1 percent or above +5 percent annually on average through 2300). Specifically, according to the Maddison Project data set—one of two data sets used by Müller, Stock, and Watson (2020)—which includes country-level GDP per capita data as far back as 1500 for some countries, no country has experienced such extreme growth for such long periods.[17] In their model, those extreme tail simulated outcomes are driven by the structure of the Bayesian model with its embedded distributional assumptions rather than by the historical data used to estimate the model.

Further, the 1st and 99th percentiles of the combined distribution of long-run growth rates based on our economic growth survey are –0.6 percent

17. For example, no country in Maddison Project data has observed 100-year growth rates below –1 percent or above +3 percent. Maddison Project data are available at Clio Infra, "GDP per Capita," https://clio-infra.eu/Indicators/GDPperCapita.html.

Figure 5. Average Projected Growth Rates of GDP per Capita in OECD Countries



Sources: Authors' calculations and Müller, Stock, and Watson (2020).

Note: Shaded areas and dashed/dotted lines represent 5th to 95th (darker, dashed) and 1st to 99th (lighter, dotted) percentile ranges.

and +4.4 percent, indicating that long-run growth rates are unlikely to fall outside this range. For these reasons, and in consultation with James Stock, we omit some projections in the extreme tails of Müller, Stock, and Watson's (2020) distribution that are outside the range of historical experience and also outside the long-run range implied by our survey (see online appendix for our approach).

Our survey provides quantiles of economic growth for the OECD for four discrete years. To maintain the rich country-level information of the econometric model while incorporating the information from the experts, we reweight the probability of occurrence of each of the 2,000 draws from Müller, Stock, and Watson (2020) to satisfy the experts' combined distribution over the long run. The underlying projections from Müller, Stock, and Watson (2020) remain unchanged (aside from the omission of extreme tails described above), but the likelihood of drawing a given trajectory is modified such that the quantiles of OECD growth reflect the distribution produced by the survey.

15985-04a_Rennert-4thPgs.indd 242 5/12/22 12:08 PM

We accomplish this reweighting in two steps. First, we generate a set of target quantiles for the years 2030, 2050, 2100, 2200, and 2300 by calculating weighted averages of the combined cumulative distribution functions from the experts and the corresponding functions from the raw data in Müller, Stock, and Watson (2020). NASEM (2017) recommended giving expert judgment increasing weight for longer horizons, so the near-term weighting is governed more by historical evidence and that of the long-term future more by the experts. For this reason, we increase the weight of the survey quantiles versus Müller, Stock, and Watson's (2020) quantiles linearly over time from zero percent in 2030 to 100 percent in 2200 and thereafter.

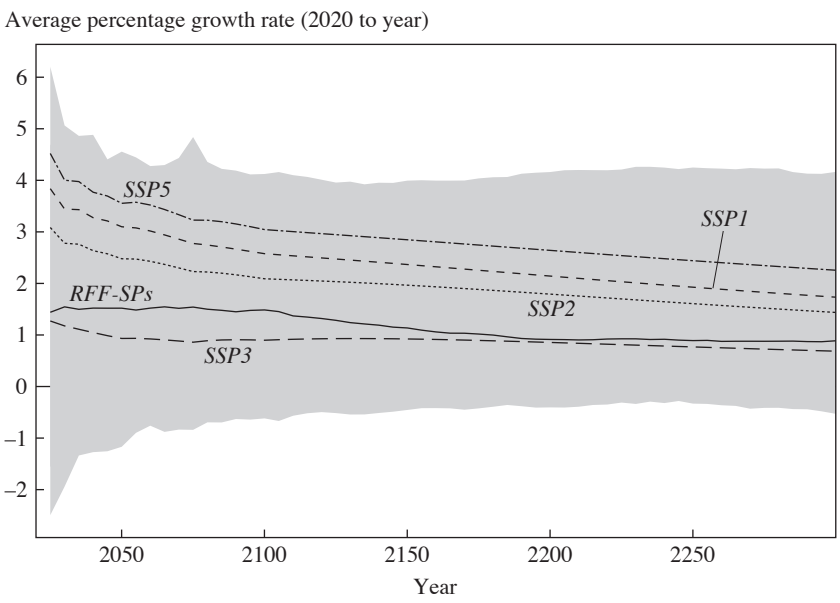
We then use iterative proportional fitting (Csiszar 1975) to impose the target quantiles for OECD growth on the 2,000 trajectories of the frontier from Müller, Stock, and Watson (2020) for each of the four benchmark years. For each range of values between each elicited quantile, this algorithm reassigns probabilities to each trajectory whose value falls within that range by minimizing a penalty for nonequal weights, subject to matching the target quantiles. Because there are four years for which we have a combined expert distribution to satisfy, the algorithm iterates between each year until all years' distributions are satisfied. Figure 5 compares the resulting distributions from Müller, Stock, and Watson (2020) with those reweighted according to our economic growth survey.

We next generate a distribution of projected global GDP per capita rates by taking 10,000 independent samples from the population and survey projections, taking the product of population and GDP per capita at the country level, summing to yield global GDP, and dividing by the global population for that draw.[18] Figure 6 shows that the resulting median global GDP growth rates from the RFF-SPs track slightly higher than SSP3, with SSP1, SSP2, and SSP5 also falling within the 90th percentile range. The SSPs do not span the full range of potential growth paths, especially below the median for

18. The raw data set from Müller, Stock, and Watson (2020) provides growth projections for 113 countries. Here we expand on that scope of coverage to include all 184 countries represented in the SSPs by undertaking the following steps to impute each country omitted in Müller, Stock, and Watson (2020): (1) identify the country within the same continent and within 30 degrees latitude with the closest matching log(GDP/capita) for the year 2020 (or, for eleven countries missing data for 2020, we use the most recent year available, typically 2019); (2) calculate a scaling factor based on the ratio between the respective 2020 GDP/capita values; and (3) apply the scaling factor to each trajectory for the matched country to generate corresponding trajectories for the omitted country. Matches for omitted countries from Oceania were identified from within Asia. The countries imputed represent a total of 3 percent of global GDP for the year used for the match.

15985-04a_Rennert-4thPgs.indd 243 5/12/22 12:08 PM

Figure 6. Average Projected Growth Rates of Global GDP per Capita



Source: Authors' calculations.

Note: The solid line represents the median value, and dark and light shading represent the 5th to 95th (darker) and 1st to 99th (lighter) percentile ranges of the RFF-SPs.

the RFF-SP growth trajectories. As will be discussed in section IV, these relatively low-growth potential paths contribute substantially to the SCC.

I.D. Projected Emissions to 2300 Based on Economic Growth: Future Emissions Survey

METHODS To generate very long run distributions of global emissions of CO_2, CH_4, and N_2O, the RFF Future Emissions Survey elicited ten experts in socioeconomic projections and climate policy who were nominated by their peers or by members of the RFF Scientific Advisory Board. The experts surveyed were based at universities, nonprofit research institutions, and multilateral international organizations. They have expertise in and have undertaken long-term projections of the energy-economic system under a substantial range of climate change mitigation scenarios.

Like our economic growth survey, the future emissions survey employed the classical model of structured expert judgment: experts first quantified their uncertainty about variables for which true values were known, for calibration and performance weighting. Experts next provided quantiles

of uncertainty (minimum, 5th, 50th, 95th, maximum, as well as additional percentiles at the expert's discretion) for four variables for a case we called Evolving Policies, which incorporates views about changes in technology, fuel use, and other conditions and is consistent with the expert's views on the evolution of future policy. The Evolving Policies case corresponds to the US federal government's approach to benefit-cost analysis, which evaluates US regulations as incremental against a more expansive backdrop of other policies and conditions and is responsive to NASEM recommendations for including future background policy in the uncertain distributions of socioeconomic projections.

Experts provided quantiles of uncertainty for (1) fossil fuel and process-related CO_2 emissions, (2) changes in natural CO_2 stocks and negative emissions technologies, (3) CH_4, and (4) N_2O for five benchmark years: 2050, 2100, 2150, 2200, and 2300. For category 1, they were also asked to indicate the sensitivity of emissions to five GDP per capita trajectories.[19]

For each expert we generate a set of cumulative distribution functions, one for each benchmark year, emissions source, and economic growth trajectory, by piecewise linear interpolation between the quantiles provided. Then, as in the economic growth survey, we generate a corresponding set of combined equal-weight cumulative distribution functions by averaging the functions in equal measure, and a set of performance-weighted cumulative distribution functions by averaging in accordance with the experts' relative performance on the calibration questions. Quantile values from the combined functions were linearly interpolated in time between each of the benchmark years to yield a distribution of piecewise linear, nonoverlapping trajectories for each emissions source and sink.

Based on the future emissions survey, we developed a distribution of emissions scenarios to pair, one to one, with our economic growth scenarios. First, we sampled from one of 10,000 economic growth trajectories, described above. Second, we sampled a value (q) on the continuous interval [0,1] to determine the percentile of the expert's emissions trajectory to evaluate. Third, at five-year intervals from 2025 to 2300 we generated an interpolated value of the qth percentile of emissions based on the realized GDP level corresponding to that GDP trajectory in that year and the qth percentile of the experts' emissions distributions for the bounding GDP values elicited. Net emissions of CO_2 were generated by sampling independent q values for direct emissions (category 1) and natural carbon stocks and negative

19. See online appendix for a more-detailed discussion of the survey methodology and the full elicitation protocol.

emissions technologies (category 2) and summing the resulting trajectories, thereby including the possibility of net negative emissions.[20]

RESULTS OF THE FUTURE EMISSIONS SURVEY Experts' performance on the calibration questions was high, as measured by statistical accuracy, informativeness, and robustness of results (see online appendix). Experts described their rationale and the conditions supporting their distributions of emissions, often citing the same factors. For direct CO_2 emissions (category 1), experts viewed low economic growth as likely to reduce emissions overall but also lead to reduced global ambition in climate policy and slower progress to decarbonization. For median economic growth conditions, experts generally viewed policy and technology evolution as the primary driver of their emissions distributions, often offering a median estimate indicating reductions from current levels but with a wide range of uncertainty. Several experts said high economic growth would increase emissions through at least 2050, most likely followed by rapid and complete decarbonization, but with a small chance of substantial continued increases in emissions. In general, the distributions were inconsistent with keeping global temperature increases below 1.5 degrees Celsius, even when considering the potential for negative emissions.

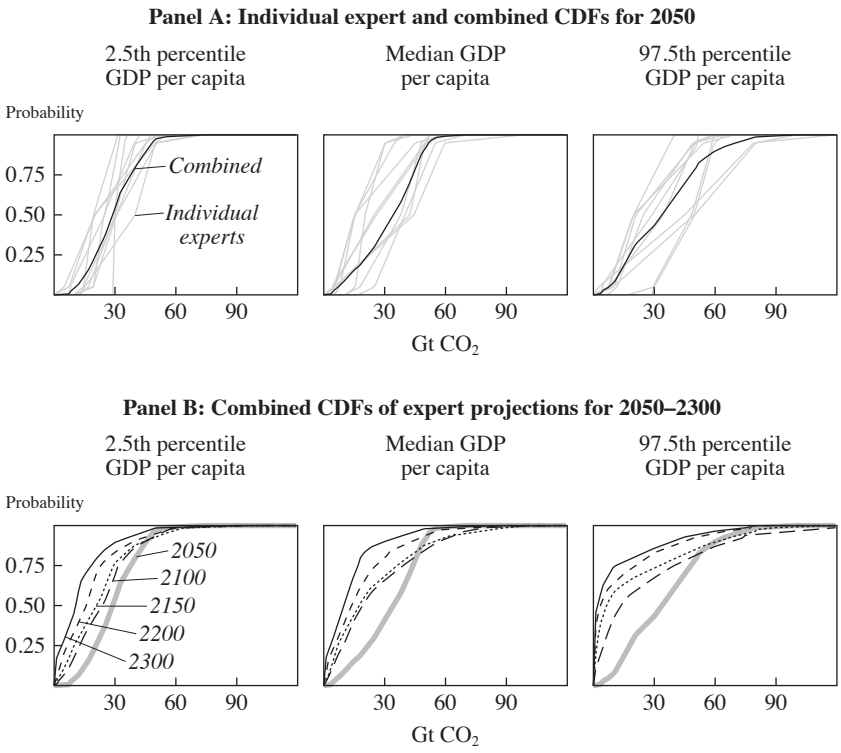
Though their rationales were often similar, experts' interpretation of those narratives, as shown in their quantiles of emissions, differed substantially (figure 7). For example, for the median growth trajectory to 2050, the median emissions ranged from 15 to 45 Gt CO_2, a span encompassing a decrease of more than 50 percent to an increase of more than 30 percent from today's

20. The experts received real-time feedback about the implications of their prescribed distributions for future outcomes. After each had provided a full set of quantiles, we followed the same sampling process described above to generate distributions of emissions trajectories, except that the emissions distributions were based on input provided by only that expert rather than the full set of experts and that for expediency we presented results based on 100 to 1,000 samples at the discretion of the expert. Experts were shown their full distributions of emissions trajectories, the economic growth paths sampled, population, emissions intensity, and the resulting climate outcomes from the FaIR 2.0 climate model (described in section II) for their verification. They were permitted to modify their quantiles after seeing their distributions and resulting climate outputs, but in general they found the results to be in agreement with the intent of their quantiles and consistent with their supporting rationale.

For each emissions trajectory generated, we used a cubic spline to interpolate between 2020 emissions and 2050 (the first quantiles provided by the experts) based on the slope of the global emissions trajectories over the 2010–2020 period and the emissions trajectory post-2050. We also used a cubic spline to interpolate trajectories between the additional years for which quantiles were provided by the experts.

15985-04a_Rennert-4thPgs.indd 246 5/12/22 12:08 PM

Figure 7. Cumulative Distribution Functions (CDFs) of Annual CO_2 Emissions

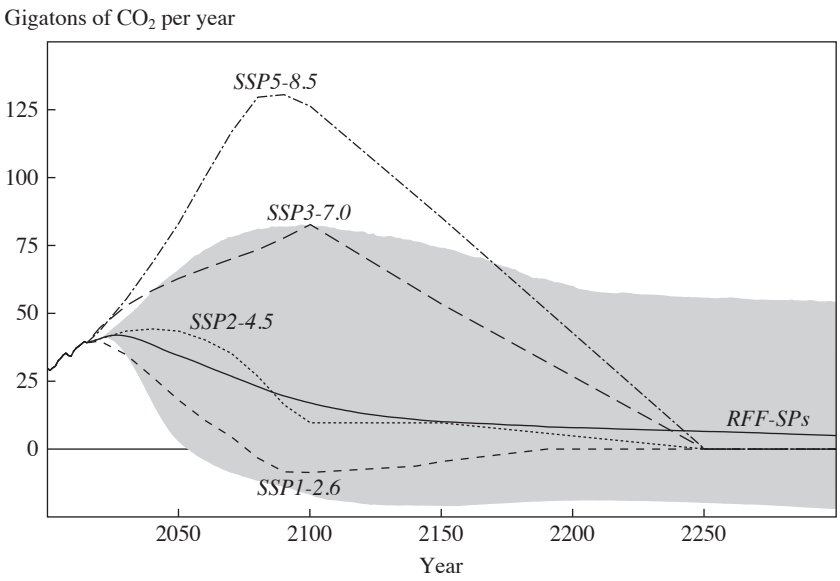


Source: Authors' calculations.

levels. Experts often provided highly skewed distributions, with significant chances that direct CO_2 emissions (category 1) would be exactly or near zero while allowing for much higher emissions in the middle and upper quantiles of their distribution.

The experts' narratives support an evolution of the combined distributions. Over time, emissions distributions for all growth trajectories exhibit a shift, particularly evident for the median and high-growth trajectories, with median emissions approaching zero in and after 2150. Emissions distributions for the lower-growth trajectory show a decreased range of emissions overall compared with the higher-growth trajectories, but the temporal trend toward lower emissions is not as strong. Higher-growth trajectories show relatively greater probabilities of increased emissions in the near term, followed by greater chances of full decarbonization in the

Figure 8. Net Annual Emissions of CO_2 from RFF-SPs and SSPs



Source: Authors' calculations.

Note: Lines represent median values, and dark and light shading represent the 5th to 95th (darker) and 1st to 99th (lighter) percentile ranges of the RFF-SPs.

next century, while also allowing for the possibility of much higher emissions over the long term.[21]

RESULTING GLOBAL GREENHOUSE GAS EMISSIONS PROJECTIONS Figure 8 shows the resulting distribution of projected net CO_2 emissions based on the future emissions survey. The median emissions trajectory is a roughly 50 percent decrease from today's levels by 2100, followed by slowly decreasing levels that approach but do not reach net zero. The median of our CO_2 emissions and concentrations paths is similar to SSP2, and the 98 percent confidence interval spans a range similar to that of SSP1 through SSP3, at least through 2140.[22] The magnitude of CO_2 emissions associated with SSP5, however, is considerably higher than the upper end (99th percentile) of our distribution through the middle of the next century, consistent with the findings of

21. See online appendix for results for all emission source and sink categories and additional discussion of the experts' rationales across all categories.

22. For comparison of emissions consistent with the SSPs beyond 2100, we adopt the commonly used extensions provided by the Reduced Complexity Model Intercomparison Project (Nicholls and others 2020).

15985-04a_Rennert-4thPgs.indd 248 5/12/22 12:08 PM

Raftery and others (2017) and Liu and Raftery (2021). Beyond the middle of the next century, all the SSP emissions trajectories increasingly lie well within our distribution because their extension beyond 2100 is constructed to achieve zero emissions by 2250. This is a weakness of the SSPs as a basis for SCC estimation, even if a subset of the SSPs spans a "reasonable range" during this century.

For CH_4 (figure OA-9 in the online appendix), the emissions distribution resulting from the future emissions survey is centered between SSP2 and SSP5 and spans a range similar to that of SSP1–SSP5, at least through 2100. After that point, as with CO_2, the emissions range spanned by the SSPs narrows, whereas the CH_4 emissions from the survey maintain a relatively wide distribution, similar to that in 2100. For N_2O (online appendix figure OA-10), the median of the emissions paths is between SSP2 and SSP5 through roughly 2200, and the full distribution from the survey spans a range wider than all the SSPs.

In sum, no single SSP is centered similarly to the median emissions paths across all three major greenhouse gases. The full range of emissions represented by the SSPs is higher than for the future emissions survey for CO_2 through 2140; by construction the range narrows to zero for CO_2 after that point and is narrower than the survey results for both CH_4 and N_2O for nearly the full period.

II. From Emissions to Monetized Climate Damages

II.A. Climate System Methods

The second step in estimating the SCC is using a climate model to calculate changes in the climate system corresponding to changes in greenhouse gas emissions. Climate models vary in their representation of the underlying physics, in their spatial and temporal resolution, and in their computational requirements. Earth system models, such as those used for IPCC analyses, require supercomputers, but SCC calculations, typically generated from tens to hundreds of thousands of samples to characterize their uncertainty, preclude use of full-scale earth system models. SCC models are designed to emulate the response of full earth system models across a subset of relevant climate outputs, such as globally averaged surface temperature.

Previous SCC calculations from the federal government used three integrated assessment models: DICE, the Climate Framework for Uncertainty, Negotiation and Distribution (FUND), and Policy Analysis of the Greenhouse Effect (PAGE), each of which employs its own reduced-form

climate model. These integrated assessment models can deliver substantially different temperature increases for the same pulse of emissions (Rose and others 2014), leading to inconsistency when results are averaged to calculate the SCC. The NASEM report therefore recommended adopting a uniform climate model that met certain criteria, including that it generates a distribution of outputs across key climate metrics comparable to distributions of outputs from the full earth system models.

The Finite Amplitude Impulse Response (FaIR) model (Millar and others 2017) was highlighted in the NASEM report as a reduced-form model that met the criteria. To assess the changes in global mean surface temperatures resulting from the RFF-SPs, we ran the latest version, FaIR 2.0 (Leach and others 2021), using 10,000 draws from the emissions trajectories of CO_2, CH_4, and N_2O while also sampling across FaIR's native uncertainty in climate variables.[23]

II.B. Resulting Temperature Change from RFF-SPs

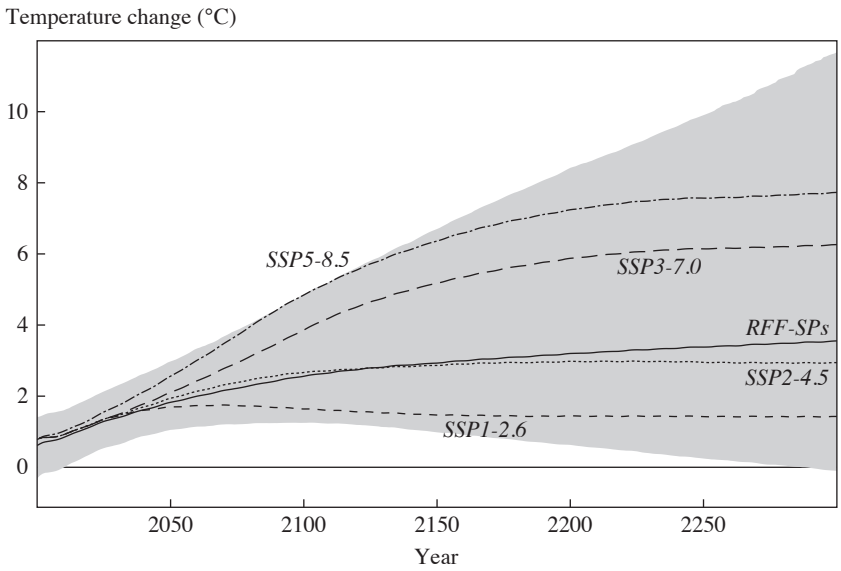
Figure 9 shows the median temperature trajectory associated with the RFF-SPs: increases reaching nearly 2.6 degrees Celsius above the average global temperature for 1850–1900 (the standard IPCC preindustrial benchmark) through 2100 and continued increases through 2300. The low end of the distribution indicates a roughly 20 percent chance that the increase will remain below 2 degrees Celsius through 2100. Our experts' expectations for negative emissions technologies lead to an increasing chance of drawing down atmospheric CO_2 to yield temperatures at current levels and below by the late 2100s.

The RFF-SP median temperature trajectory tracks closely with SSP2 through 2150, thereafter continuing to increase slightly. SSP1 is largely consistent with the 5th percentile results throughout the period. Temperatures resulting from SSP3 emissions are consistent with the 95th percentile of the RFF-SPs through the middle of the next century, at which point temperatures stop increasing, by construction. The median temperature from SSP5 is roughly consistent with the 99th percentile of temperatures from the RFF-SPs through 2100, at which point it begins to level off to meet the imposed requirement for net zero emissions by 2250.

In this comparison, uncertainty in the climate system itself, as represented by the uncertain distributions of climate parameters in the FaIR model, contributes significant uncertainty to the range of projected temperatures. The temperature distributions for the RFF-SPs include climate uncertainty

23. Trajectories for non-CO_2, CH_4, and N_2O were drawn from SSP2.

Figure 9. Global Mean Surface Temperature Change from RFF-SPs and SSPs



Source: Authors' calculations.

Note: Temperature change is relative to the standard 1850–1900 preindustrial average. Solid lines represent median values. Dark and light shading represent the 5th to 95th (darker) and 1st to 99th (lighter) percentile ranges based on the RFF-SPs. For clarity of presentation, uncertainty in the climate system is reflected in the uncertainty range only for the RFF-SPs (and not the SSPs).

from FaIR, but for clarity we omit climate system uncertainty in presenting projected temperatures from the SSPs. For a sense of scale, the 90th percentile range in temperatures from FaIR in 2300 for SSP5 is about –2.5 to +7 degrees Celsius about the median.

METHODS FOR CLIMATE DAMAGE ESTIMATION The third step in estimating the SCC is translating changes in the climate system, such as temperature, into total economic damages over time. Damages can be calculated by estimating costs for various sectors (e.g., human health and mortality, agriculture, energy usage, coastal flooding) and summing them, or by taking an aggregate approach to estimate damages across the economy as a whole.

Recent advances in methodologies for damage estimation are not reflected in the integrated assessment models used by the federal government to calculate the SCC (NASEM 2017; Diaz and Moore 2017). The NASEM report made recommendations on improving sectoral damage estimation, finding sufficient peer-reviewed research to support updates on human health and mortality, agriculture, coastal inundation, and energy

demand. Since the report was issued, the literature addressing specific sectors has grown.

Nevertheless, few studies meet the full requirements (e.g., global coverage with regional detail, translation into economic damages) put forward by Diaz and Moore (2017) or Raimi (2021) to serve as the basis for an updated damage function for the SCC. For example, two independent, comprehensive reviews (Bressler 2021; Raimi 2021) found just three suitable studies (World Health Organization 2014; Gasparrini and others 2017; Carleton and others 2018). Our own further assessment of the damages literature found two candidates for agricultural damages (Moore and others 2017; Calvin and others 2020), two for energy demand (Clarke and others 2018; Ashwin and others 2021), and one for coastal damages (Diaz 2016).

Among the notable additions, the Climate Impact Lab has developed a methodology to generate empirically derived, hyper-localized damage functions accounting for adaptation. The Climate Impact Lab in its research has been applying its methodology across a comprehensive set of sectors including health, agriculture, labor, energy, conflict, coastal, and migration (Carleton and others 2018). Upon completion, this full set of sectors is intended to support fully empirically based climate damage estimates.

Much of the new sectoral damages research identified here is currently under peer review for publication, and efforts to implement the existing peer-reviewed studies will similarly be completed on a timeline that is compatible with the IWG process to update the SCC. As described below, for the purposes of this paper we have deployed the aggregate global climate damage function from the widely used DICE model (Nordhaus 2017b) to develop illustrative SCC estimates, coupled with the RFF-SPs, the FaIR climate model, and the stochastic discounting approach described in the next section.

III. Discounting Approaches for the Social Cost of Greenhouse Gases

The long residence time of CO_2 in the atmosphere implies that today's emissions will have consequences for centuries. This time horizon makes the discount rate a major factor for the SCC. For example, the IWG's 2021 interim SCC estimate is $51/ton with a 3 percent discount rate (IWG 2021) but would be about $121/ton at a 2 percent discount rate (RFF and NYSERDA 2021). That 1 percentage point difference alone would more than double the SCC and, by implication, greatly strengthen the economic rationale for substantial emissions reductions.

The discount rates used in federal regulatory analysis are guided by Circular A-4, issued by the Office of Management and Budget (OMB) in 2003, which endorses rates of 3 percent and 7 percent reflecting, respectively, consumption and investment rates of return (White House 2003). OMB guidance also allows for additional sensitivity analysis in cases with intergenerational consequences, such as climate change. However, this guidance runs counter to current economic thought and evidence, for three reasons: (1) a constant deterministic discount rate becomes increasingly problematic for long-horizon problems (Weitzman 1998); (2) benchmarks for the consumption rate of interest (currently 3 percent) have declined substantially over the past two decades (CEA 2017; Bauer and Rudebusch 2020, 2021); and (3) the rationale for 7 percent—to address possible policy effects on capital—is flawed in ways that are magnified for very long term decisions (Li and Pizer 2021).

The NASEM (2017) report and recent technical guidance on the SCC (IWG 2021) acknowledged those concerns. A 2021 executive order directed the OMB to reassess existing practice and consider "the interests of future generations" in revisions to Circular A-4 (White House 2021a, sec. 2). Alongside issues related to empirical discount rate uncertainty over long time horizons, the comparison of welfare across generations creates an ethical concern dating back at least as far as Ramsey (1928): Do we discount the welfare of future generations simply because they are born later?

One rationale for changing the government's discounting approach is the systemic decline in observed interest rates over at least the past two decades (Kiley 2020; Del Negro and others 2017; Johannsen and Mertens 2016; Laubach and Williams 2016; Caballero, Farhi, and Gourinchas 2017; Christensen and Rudebusch 2019; CEA 2017; Rachel and Summers 2019; Bauer and Rudebusch 2020, 2021), which along with other research on discount rates for very long-run horizons (Giglio, Maggiori, and Stroebel 2015; Giglio and others 2021; Drupp and others 2018; Carleton and Greenstone 2021) has led to calls for using a lower discount rate; 2 percent is often suggested.

The second argument for a modified discounting approach stems from uncertainty in the discount rate, which tends to lead to declining future discount rates. Weitzman (1998) showed that if one is uncertain about the future trajectory of (risk-free) discount rates, and uncertain shocks to the discount rate are persistent, the certainty-equivalent (risk-free) discount rate declines with the time horizon toward the lowest possible rate. This result stems from a straightforward application of Jensen's inequality to a stochastic discount factor, leading to declining (risk-free) discount rates

(Arrow and others 2014). At the same time, if the payoffs to investments in emissions reductions are correlated with future income, the effective *risk-adjusted* rate could be higher if the correlation is positive or lower if it is negative (Gollier 2014). This correlation is often termed the "climate beta," but it is not clear ex ante whether the beta is positive, as in Nordhaus's work and as argued by Dietz, Gollier, and Kessler (2018), or negative, as in Lemoine (2021).

The third issue is the need, in light of recent research (Li and Pizer 2021), to rethink the use of the higher discount rate (7 percent) reflecting the return to capital. Several decades ago, researchers suggested that when taxes create a wedge between consumption and investment interest rates, the alternative rates could be used to bound a benefit-cost analysis, as a shorthand version of the shadow price of capital (SPC) approach (Harberger 1972; Sandmo and Drèze 1971; Marglin 1963a, 1963b; Drèze 1974; Sjaastad and Wisecarver 1977). However, the assumptions underlying the soundness of that approach are quite restrictive: costs are assumed to occur entirely in the first period; benefits are constant and occur either in a single period or in perpetuity; and benefits displace only consumption while costs displace either investment or consumption. Li and Pizer (2021) extend Bradford (1975), showing that the traditional approach of using 7 percent as a shorthand means for representing investment impacts of regulatory costs becomes increasingly very inaccurate the farther one looks into the future.

The NASEM (2017) report foreshadowed those results and recommended using a central consumption rate estimate along with sensitivity cases. Newell, Pizer, and Prest (2021) provide some guidance, examining central values of 2 percent and 3 percent and a range of values between 1.5 percent and 5 percent (though they do not recommend those particular values). Their discussion of discount rates is based primarily on questions about the most appropriate near-term consumption rate and does not address the SPC approach. Pizer (2021) details how the SPC approach could be implemented, suggesting sensitivity cases that employ the consumption discount rate, with costs and benefits alternately multiplied by the SPC to reflect the possibility that the entirety of each of these impact streams falls on investment: an SPC of 1.2 is proposed as a conservative value. Alternatively, simply multiplying regulatory costs by the SPC provides a sensitivity case that is consistent with an (extreme) scenario where all costs fall on investment. Conceptually, this is equivalent to what is being sought with the traditional approach of discounting benefits at the higher 7 percent rate, but it has the advantage of both being analytically correct and allowing for a

15985-04a_Rennert-4thPgs.indd 254 5/12/22 12:08 PM

consistent discounting approach across different elements of benefit-cost analysis. The consumption discount rate would be employed in all cases, and the SPC approach would apply generally, not just in the context of the SCC.

Each of these discounting ideas (including stochastic growth discounting, discussed below) could be incorporated in a revision to Circular A-4, with relevance to both SCC estimation and other contexts. This would harmonize SCC discounting and broader US government guidance on benefit-cost analysis.

III.A. Stochastic Growth Discounting with Economic Uncertainty

One rationale for discounting, generally, is the concept of declining marginal utility of consumption. Intuitively, a $100 cost in a future in which society has grown dramatically wealthier should be valued less, from today's perspective, than the same $100 cost in a relatively poor future with stagnant economic growth. This result is often embodied by the classic equation derived in Ramsey (1928) that relates the consumption discount rate (r_t) to the rate of consumption growth (g_t) over time:

$$r_t = \rho + \eta g_t. \tag{1}$$

In equation (1), ρ represents the rate of pure time preference (how much utility is discounted over time) and η represents the curvature of an isoelastic utility function.[24] We use time subscripts to refer to the compound average value of the indicated variable from today (time 0) to year t. If average consumption growth to year t, g_t, is uncertain, as it is given the probabilistic socioeconomic scenarios discussed earlier, then the average discount rate to year t, r_t, is also uncertain. This leads to a stochastic discount factor, which is used to discount stochastic marginal damages from an incremental ton of emissions (MD_t) to a present value (PV) equivalent:

$$PV\left(MD_t\right) = E\left[e^{-r_t t} MD_t\right], \tag{2}$$

where r_t is determined by equation (1) based on the uncertain growth rate g_t. An alternative is to base the discount rate on some market proxy for the discount rate as in Bauer and Rudebusch (2021). Either way, the discount rate is considered uncertain, and the first term inside the expectation, $e^{-r_t t}$,

24. $u(c) = c^{1-\eta}/(1-\eta)$.

represents a stochastic discount factor. In our treatment, the discount factor and rate are uncertain due to the stochastic growth rate. The importance of a stochastic discount factor is well established in the finance literature, and its importance is increasingly recognized in the literature at the nexus of macro and climate economics (Cai and Lontzek 2019; Barnett, Brock, and Hansen 2020, 2021). A stochastic discount rate leads to a declining certainty-equivalent risk-free rate (Weitzman 1998). To clearly see the derivation of this result, suppose for the moment that the discount rate is normally distributed, $r_t \sim N(\mu_t, \sigma^2)$, and that it is uncorrelated with marginal damages, $corr(e^{-r_t t}, MD_t) = 0$, which corresponds to a climate beta of zero. Then it is easy to show that the certainty-equivalent rate, which is denoted r_t^{ce} and represents the rate at which to discount expected marginal damages (as in $e^{-r_t^{ce} t}E[MD_t]$), declines with the time horizon of the impacts being discounted, t:[25]

$$r_t^{ce} = \mu_t - \frac{1}{2}t\sigma^2. \tag{3}$$

Of course, equation (3) represents a special case. More generally, absent these two specific assumptions, the risk-free rate given by this equation does not account for the risk profile of the benefits of emissions reductions, namely, through the climate beta, which reflects the potential correlation of the stochastic discount rate with marginal damages. If one wants to retain the certainty-equivalent approach to discounting, Gollier (2014) shows that a risk adjustment is necessary to account for any such correlation, but the form of this adjustment depends on the potentially complex nature of the joint uncertainties. We instead take a more general approach to account for these issues by directly using the more general equations (1) and (2) to implement stochastic discounting as part of the Monte Carlo estimation of the SCC, which explicitly accounts for any such correlation. Accounting for this correlation is important in theory (Barnett, Brock, and Hansen 2021) and also,

25. A version of this result is shown in Newell and Pizer (2003), but for clarity of exposition we explain it briefly here. Starting with the definition that the certainty-equivalent rate yields the same present value of equation (2), we have $e^{-r_t^{ce} t}E[MD_t] = E[e^{-r_t t}MD_t] = E[e^{-r_t t}]\ E[MD_t]$, where the last equality follows by the assumption of zero correlation. Solving for r_t^{ce} yields $r_t^{ce} = \frac{1}{-t}\log(E[e^{-r_t t}])$. A well-known property of the exponential function, e^x, applied to a normally distributed variable, $x \sim N(\mu, \sigma^2)$, is that $E[e^{ax}] = e^{a\mu + \frac{1}{2}a^2\sigma^2}$. Applying this formula with $x = r_t$ and $a = -t$ yields the result.

as our results show, matters greatly in practice when the climate beta is not zero. Indeed, the climate beta in most integrated assessment models is implicitly taken to be close to one.

For example, in the DICE model, damages are assumed to be a percentage of GDP (where that percentage depends on global temperature), and the discount rate is a linear function of economic growth, as in a Ramsey-like framework (Nordhaus and Sztorc 2013). This implies a beta of essentially one, since higher income (and, in turn, greater discounting) is perfectly correlated with higher undiscounted damages. That is, a positive beta implies that undiscounted damages are largest when economic growth is largest, and smallest when growth is smallest. Mirroring this, with $\eta > 0$, the discount factor is smallest when growth is largest and largest when growth is smallest. Using a stochastic discount factor as in equation (2) will therefore discount damages most in states of the world where they accrue to rich future generations and correspondingly discount them least in states where the future is poor. Amid uncertainty about socioeconomic trajectories, ignoring this stochastic discount factor (and its correlation with climate impacts) could severely bias estimates of the SCC. The magnitude of this bias depends on the climate beta and on the nature of the uncertainty in socioeconomic and emissions trajectories; Newell, Pizer, and Prest (2021) and our illustrative results (below) show that this bias could change the SCC by a factor of two or more.

Despite the importance of stochastic discounting, federal government benefit-cost analysis has historically not treated the discount rate as explicitly uncertain, nor has the discount rate been connected to growth as in the Ramsey framework. Instead, the consumption discount rate used in past government estimates of the SCC has been a constant rate of 3 percent.[26] This is equivalent to implicitly choosing the discounting parameters $\rho =$ 3 percent and $\eta = 0$, corresponding to a linear utility function. Yet this approach effectively eliminates any consideration of declining discount rates, as in Weitzman (1998), and risk premia, as in Gollier (2014). More intuitively, it also treats a $100 cost to a member of a wildly rich future generation the same as a $100 cost to a poor one, which is incorrect from

26. Although 3 percent was the central rate, the IWG also previously used constant rates of 2.5 and 5 percent as sensitivity cases. Because those values were estimated to roughly approximate the effects of explicitly accounting for uncertainty in risk-free and risk-adjusted rates (Newell, Pizer, and Prest 2021), those motivations are no longer appropriate when stochastic discounting can be captured explicitly in integrated assessment models, as we propose.

a welfare perspective. Correspondingly, such parameter values receive little support from economists working in this field (Drupp and others 2018).

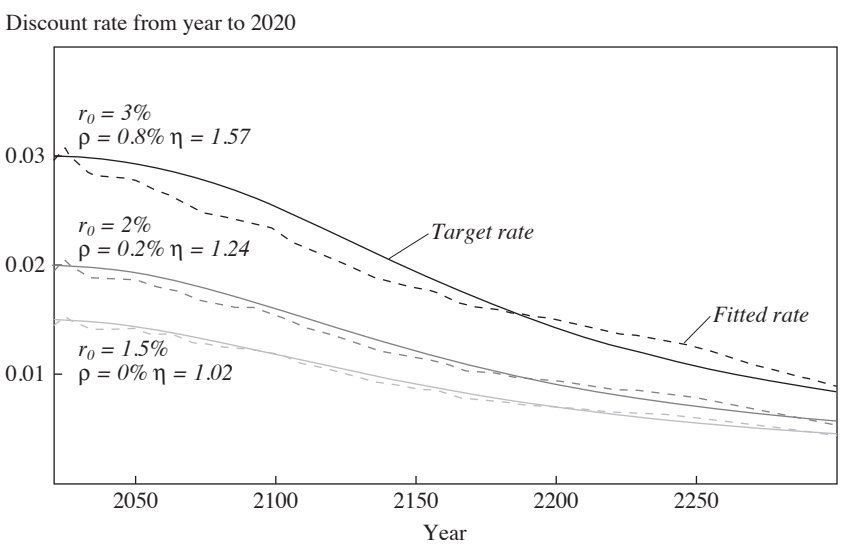
Although the case for using stochastic discounting as in equation (1) is strong, the choice of the parameters in that equation is not a simple matter, and their values can lead to very different effective discount rates (Stern 2007; Nordhaus 2017a) and their connection to economic growth and climate damages. One recent paper surveyed economists about their preferred values of ρ and η (Drupp and others 2018). This is valuable, but the federal government has a long tradition of relying on descriptive, empirical approaches to informing discounting guidance, as in other aspects of benefit-cost analysis. In particular, Circular A-4 refers to observed interest rates in selecting 3 and 7 percent (White House 2003). A choice of ρ and η might therefore sensibly start with the constraint that the associated near-term rate match the consumption rate used elsewhere in benefit-cost analysis, as recommended in NASEM (2017). However, a continuum of (ρ, η) combinations can match any particular near-term rate, so another constraint is needed.

Newell, Pizer, and Prest (2021) provide such an approach. They calibrate the values of (ρ,η) such that, when applied to the Müller, Stock, and Watson (2020) growth distribution, the implied discount rate term structure starts at a specified rate in the near term (say, 3 or 2 percent) before declining with the time horizon in a manner consistent with evidence from the empirical literature on future interest rate term structures (Bauer and Rudebusch 2020, 2021).[27] Figure 10 illustrates the calibrated combinations of (ρ, η) yielding implied (fitted) term structures when applied to the RFF-SPs (dashed lines). These parameters were calibrated to be as consistent as possible with those implied by the Bauer and Rudebusch (2021) model initialized to given targeted near-term rates of 1.5, 2, or 3 percent (solid lines). For example, using the estimated model from Bauer and Rudebusch (2021) and starting with a near-term rate of 2 percent, we construct a target term structure (solid black curve). We then find the combination, (ρ, η) = (0.2 percent, 1.24), that best fits the target term structure.

27. The parameters shown here differ slightly from those in Newell, Pizer, and Prest (2021) because we calibrate them to the full RFF-SPs, corresponding to the Müller, Stock, and Watson (2020) distribution weighted based on our economic growth survey. The methodology developed in Newell, Pizer, and Prest (2021) was demonstrated on the raw distribution, before the weights were applied.

15985-04a_Rennert-4thPgs.indd 258 5/12/22 12:08 PM

Figure 10. Calibrated Certainty-Equivalent Risk-Free Term Structures and Target Term Structure



Source: Authors' calculations based on Bauer and Rudebusch (2021) and Müller, Stock, and Watson (2020).

The calibration procedure in Newell, Pizer, and Prest (2021) can be implemented for any specified near-term rate. Here we present three cases: 1.5 percent, 2 percent, and 3 percent:[28]

$$r_t^{1.5\%} = 0\% + 1.02 g_t$$

$$r_t^{2\%} = 0.2\% + 1.24 g_t$$

$$r_t^{3\%} = 0.8\% + 1.57 g_t$$

These ρ and η parameters lie in the middle of the range often used in the literature, particularly for target near-term rates of 3 percent and 2 percent. Implementing them simultaneously with the socioeconomic trajectories discussed in section I produces a declining term structure of

28. See Newell, Pizer, and Prest (2021, sec. 3.2) for the rationale behind each rate, and that paper's appendix for additional alternative near-term rates.

certainty-equivalent, risk-free rates consistent with the empirical literature (Bauer and Rudebusch 2020, 2021). Importantly, implementing the stochastic discount rate alongside stochastic damages via equation (2) explicitly captures risk aversion and the correlation between the discount rate and climate damages, meaning no ex post risk adjustment to the discount rate is necessary.

This calibrated stochastic discounting rule can now be used with the undiscounted damage estimates (discussed above) to estimate the SCC in an internally consistent manner.

IV. Illustrative Calculations of the Social Cost of Carbon

We present illustrative estimates of the SCC based on our socioeconomic projections (the RFF-SPs), the FaIR climate model, and our discounting methodology—all of which speak directly to the NASEM (2017) recommendations—and apply them using the DICE damage function (Nordhaus 2017a). This approach is directly responsive to three of the four NASEM recommendations. The fourth recommendation is to update the damage functions with the best available science on sectoral damages, rather than using an aggregate damage function such as that in DICE. We will include more recent sector-specific damage estimates, reflecting the best available science, in future work, but for the moment we use the DICE damage function to produce illustrative SCC estimates. Although the values we present here should be considered illustrative, they highlight the importance of socioeconomic uncertainty and stochastic growth discounting, and the interaction of these two important drivers of the SCC.

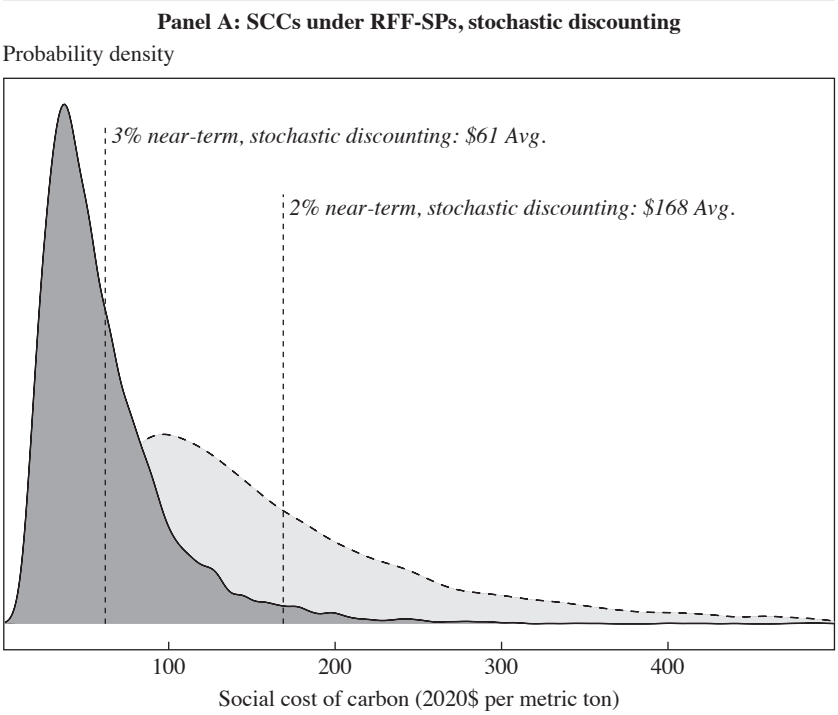
We also compare our SCC estimates with those from SSPs 1, 2, 3, and 5. Because of the lack of socioeconomic uncertainty in each SSP—and the lack of relative probabilities across them—we cannot meaningfully calibrate ρ and η parameters for those scenarios to deliver comparable near-term rates. We therefore apply constant discount rates of 2 percent and 3 percent to the SSPs.[29]

The results are shown in figure 11, leading to our first major conclusion: a quantitative probabilistic accounting of socioeconomic uncertainty matters

29. As an additional comparison, figure OA-13 in the online appendix presents an analogous figure to figure 11 but applying our RFF-SP-calibrated discounting parameters (ρ, η) to the SSPs. This is purely for presentational purposes, and we caution that our discounting parameters were not calibrated to the SSPs. Because the SSPs have no uncertainty within them, it is not possible to calibrate discounting parameters to them as we can do to the socioeconomic distributions (Newell, Pizer, and Prest 2021).

15985-04a_Rennert-4thPgs.indd 260 5/12/22 12:08 PM

Figure 11. Illustrative Probability Distributions of Social Cost of Carbon (2020$/ton CO_2)



Source: Authors' calculations.

Note: The SCC estimates should be considered illustrative because they are based on alternative socioeconomic inputs, discounting approaches, the FaIR 2.0 climate model, and the DICE damage function.

Table 1. Effects of Discounting and Growth Distribution Tails on Illustrative SCC Estimates (2020$/ton CO_2)

	Stochastic growth discounting			
	3 percent near-term $\rho = 0.8\%$ $\eta = 1.57$	*2 percent near-term* $\rho = 0.2\%$ $\eta = 1.24$	*Constant discounting*	
			3 percent constant	*2 percent constant*
Mean SCC, full distribution	$61.4	$168.4	$194	$1,557
Mean SCC, drop top and bottom 1 percent global income draws	$60.6	$167.9	$96	$450

Source: Authors' calculations.

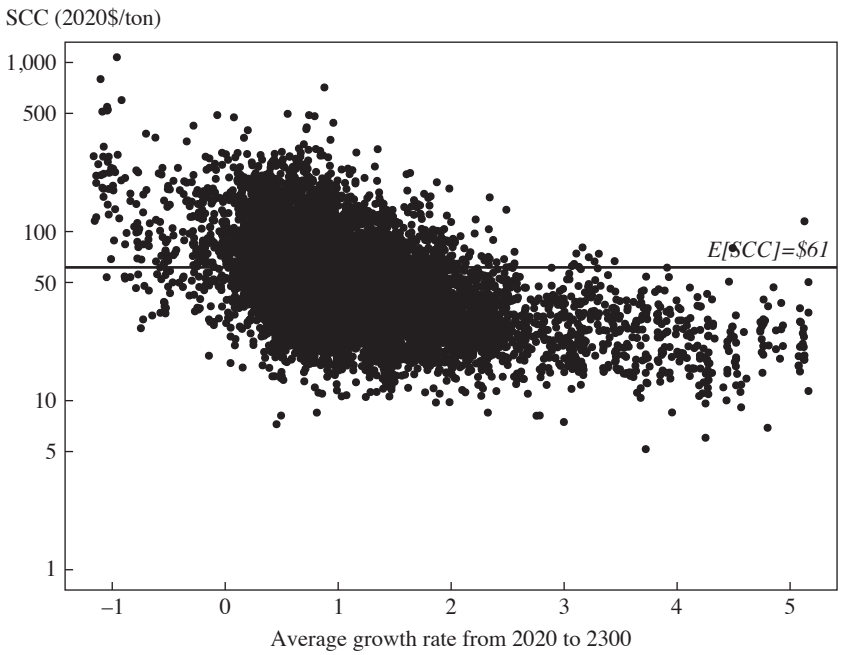
Note: The SCC estimates should be considered illustrative because they are based on alternative socioeconomic inputs, discounting approaches, the FaIR 2.0 climate model, and the DICE damage function.

greatly for the SCC. Panel A of figure 11 shows the distributions of our illustrative SCC values calibrated to 2 percent and 3 percent discount rates in the near term (means are also in the left columns of table 1). The other two panels show the SCC distributions under each SSP at 3 percent (panel B) and 2 percent (panel C) discount rates. Panel A reflects socioeconomic uncertainty implicitly, leading to central SCC estimates of $61 and $168/ton CO_2 under 3 percent and 2 percent near-term stochastic discounting, respectively. The distribution underlying those means reflects both socioeconomic and climate uncertainty. We disaggregate that distribution below, but the bottom panel shows the importance of socioeconomic uncertainty explicitly by comparing across the SSPs. SSP5 (high income growth) produces mean SCC values three to six times higher than the other SSPs. Hence, if one were to use a weighted combination of the SSPs, the resulting average SCC would reflect the relative weight given to each SSP, especially SSP5—a choice with no clear empirical basis. This result highlights the importance of incorporating a quantitative accounting of economic uncertainty, as in the RFF-SPs.

Next, figure 12 demonstrates the effect of stochastic versus constant discounting on the mean SCC, leading to our second major conclusion: stochastic growth discounting is crucially important to SCC estimation in the context of socioeconomic uncertainty. In table 1, the first two columns of the first row show mean SCCs under the RFF-SPs for stochastic growth discounting approaches consistent with 3 percent and 2 percent near-term rates (both in 2020 dollars), producing mean SCC estimates of $61.4 and $168.4/ton CO_2, respectively. These estimates, reflecting the updated socioeconomic, emission, climate, and discounting modules (three of the four

15985-04a_Rennert-4thPgs.indd 262 5/12/22 12:08 PM

Figure 12. Illustrative SCC Estimates versus Average GDP per Capita Growth Rate (2020 to 2300)



Source: Authors' calculations.

Note: The values plotted correspond to estimates of the SCC and long-run cumulative growth rates using a 3 percent near-term stochastic growth discounting ($\rho = 0.8\%, \eta = 1.57$), the RFF-SPs, the FaIR 2.0 climate model, and the DICE damage function.

NAS recommendations), are 33 percent and 50 percent higher than the corresponding DICE-only SCC estimates from the 2016 IWG ($46 and $112/ton CO_2 in 2020 dollars; RFF and NYSERDA 2021).

We also present the results from the RFF-SPs with constant discounting to illustrate the importance of stochastic discounting, but as previously discussed, constant discounting is inappropriate when uncertainty in economic growth is considered, as here. When constant discounting is coupled with uncertain growth, the mean SCC is higher than is appropriate by a factor of three to nine, $194 and $1,557/ton CO_2 for 3 percent and 2 percent discount rates, respectively, because it ignores the correlation between damages and growth (the climate beta) and hence the discount rate. In other words, ignoring the risk profile of the SCC threatens to overstate the mean SCC in this example by a factor of three or more ($194 versus

$61/ton with 3 percent discounting, and $1,557 versus $168/ton with 2 percent discounting).

More specifically, the high expected values reflect a right-skewed distribution of damages. It is well known that skewed distributions and tail events can influence the expected value of the benefits of mitigating climate change (Gollier 2008; Weitzman 2011, 2014). Under constant discounting, such tail events are very rich futures with associated large amounts of consumption at risk from climate change. Yet constant discounting treats each dollar of cost to those wealthy future generations the same as a dollar of cost to a relatively poor future. Hence, with constant discounting, the effects on the future rich inappropriately dominate the expected value of the SCC, leading to a strong upward bias in the SCC estimate.

This problem is recognized in the finance literature as the result of ignoring the risk properties of an investment—namely, the correlation of an uncertain payoff with the stochastic discount rate. Stochastic growth discounting addresses this by discounting the high-growth, high-damage states at a higher rate. By discounting high-growth states more, stochastic discounting stabilizes the mean and variance of the SCC, as documented in Newell, Pizer, and Prest (2021).

The second row of table 1 highlights this greater stability under stochastic discounting by showing a sensitivity case in which we drop the top and bottom 1 percent of the global average income trajectories.[30] Under constant discounting, the mean SCC is quite sensitive to dropping these 1 percent extremes, falling from $194 to $96/ton at a 3 percent discount rate and from $1,557 to $450/ton at a 2 percent discount rate. By contrast, the mean SCC is virtually unchanged under stochastic discounting, changing by less than 1.5 percent for each of the stochastic rates that are consistent with 2 percent and 3 percent near-term rates. More generally, the SCCs with stochastic discounting change only negligibly even when much larger percentiles are dropped from the tails. For example, with stochastic discounting, the mean SCCs also change by less than 1.5 percent even when the top and bottom 10 percent of draws of global average income trajectories are dropped, whereas under constant discounting, the mean SCCs fluctuate by factors of 3 to 11. This result highlights the stabilizing effect of properly incorporating stochastic growth discounting, as anticipated in the NASEM (2017) report.

30. Specifically, we drop the draws with global average GDP per capita in 2300 in the top 1 percent and bottom 1 percent of draws, before taking the average SCC.

15985-04a_Rennert-4thPgs.indd 264 5/12/22 12:08 PM

This stability with stochastic discounting is apparent in figure 12, which plots the individual Monte Carlo SCC draws against each draw's long-run (2020–2300) global GDP per capita growth rate, under the 3 percent near-term stochastic discounting parameters ($\rho = 0.8\%$, $\eta = 1.57$).[31] Roughly speaking, the vertical spread of SCC values in the figure largely reflects climate uncertainty for each given level of growth in GDP per capita, whereas the horizontal spread of SCC values reflects uncertainty in long-run income growth. Because the DICE damage function is proportional to GDP, undiscounted marginal damages scale roughly one-for-one with income growth, but with $\eta > 1$ they are discounted somewhat more than one-for-one with stochastic discounting, leading to a modest negative relationship between the SCC and GDP per capita growth. In other words, the SCC is higher when income growth is lower, and vice versa.

V. Conclusion

Since the SCC is a vitally important metric guiding climate policy, its calculation must be supported by the best available science, including an explicit incorporation of uncertainty. Our results demonstrate that socioeconomic uncertainty and stochastic discounting are important drivers of the SCC, and our work presents an opportunity to incorporate those uncertainties into ongoing updates.

Although the SCC estimates presented here are meant to be illustrative and use a highly simplified estimate of climate damages, they nonetheless highlight two major conclusions. First, socioeconomics matter significantly to the SCC, highlighting the importance of a quantitative accounting of socioeconomic uncertainty. Whereas scenario-based socioeconomic projections like the SSPs have no formal probabilities attached to them, our approach to quantifying uncertainties in future trajectories of population, GDP, and emissions helps account for these uncertainties in the SCC. Second, when incorporating socioeconomic uncertainty, stochastic growth discounting is crucial to account for the correlation of climate damages and the discount rate, whereas ignoring it leads to a large upward bias in the SCC estimate. Our work represents an advance in uncertain socioeconomic trajectories and discounting approaches based on empirically based explicitly probabilistic methods. Nevertheless, potentially important components

31. The shape of the curve is similar under the 2 percent near-term parameters ($\rho = 0.2\%$, $\eta = 1.24$) but shifted up to a higher level.

have not yet been fully incorporated into officially adopted SCC values. Recent work has begun to account for how the risk of tipping points influences the SCC (Dietz and others 2021). Other important factors include climate-related migration, conflict, and loss of at-risk species. Another conceptual issue is equity weighting, wherein effects on poorer regions of the world could be weighted more than equivalently sized dollar value to rich regions (Errickson and others 2021). Future research in these areas could be incorporated into official SCC values over time.

More generally, the SCC should be continually updated as the scientific frontier advances, as recommended by NASEM (2017). Our work speaks directly to those NASEM recommendations and presents an opportunity for the US government to improve on simple, deterministic approaches to socioeconomic projections and discounting methodologies to better reflect the interrelated uncertainties about future population, income, emissions, climate, and discount rates.

ACKNOWLEDGMENTS We thank Jordan Wingenroth, Lauren Dunlap, and Anagha Komaragiri for their expert assistance in preparing this manuscript. We gratefully acknowledge the time and insights of the participants in our Expert Workshop on Very Long-Term Projections of Population (Juha Alho, Leontine Alkema, Jakub Bijak, Patrick Gerland, Nico Keilman, Ronald Lee, Jim Oeppen, Warren Sanderson, and Tomáš Sobotka); Economic Growth Survey (Daron Acemoglu, Erik Brynjolfsson, Jean Chateau, Melissa Dell, Robert Gordon, Mun Ho, Chad Jones, Pietro Peretto, Lant Pritchett, and Dominique van der Mensbrugghe); and Future Emissions Survey (Joseph Aldy, Sally Benson, Geoff Blanford, Leon Clarke, Elmar Kriegler, Jennifer Faye Morris, Sergey Paltsev, Keywan Riahi, Susan Tierney, and Detlef van Vuuren). The work of Raftery and Ševčíková was supported by NIH grant R01 HD070936 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. RFF's work under the Social Cost of Carbon Initiative has been supported by the Alfred P. Sloan Foundation, the Hewlett Foundation, and individual donors.

15985-04a_Rennert-4thPgs.indd 266 5/12/22 12:08 PM

References

Alkema, Leontine, Adrian E. Raftery, Patrick Gerland, Samuel J. Clark, François Pelletier, Thomas Buettner, and Gerhard K. Heilig. 2011. "Probabilistic Projections of the Total Fertility Rate for All Countries." *Demography* 48, no. 3: 815–39. https://doi.org/10.1007/s13524-011-0040-5.

Arrow, Kenneth J., Maureen L. Cropper, Christian Gollier, Ben Groom, Geoffrey M. Heal, Richard G. Newell, and William D. Nordhaus. 2014. "Should Governments Use a Declining Discount Rate in Project Analysis?" *Review of Environmental Economics and Policy* 8, no. 2: 145–63. https://doi.org/10.1093/reep/reu008.

Ashwin, R., T. Carleton, M. Delgado, M. Greenstone, T. Houser, S. Hsiang, and others. 2021. "Estimating a Social Cost of Carbon for Global Energy Consumption." *Nature* 598:308–14. https://doi.org/10.1038/s41586-021-03883-8.

Azose, Jonathan J., and Adrian E. Raftery. 2015. "Bayesian Probabilistic Projection of International Migration." *Demography* 52, no. 5: 1627–50. https://doi.org/10.1007/s13524-015-0415-0.

Azose, Jonathan J., Hana Ševčíková, and Adrian E. Raftery. 2016. "Probabilistic Population Projections with Migration Uncertainty." *Proceedings of the National Academy of Sciences of the United States of America* 113, no. 23: 6460–65. https://doi.org/10.1073/pnas.1606119113.

Barnett, Michael, William Brock, and Lars Peter Hansen. 2020. "Pricing Uncertainty Induced by Climate Change." *Review of Financial Studies* 33, no. 3: 1024–66. https://doi.org/10.1093/rfs/hhz144.

Barnett, Michael, William Brock, and Lars Peter Hansen. 2021. "Climate Change Uncertainty Spillover in the Macroeconomy." In *NBER Macroeconomics Annual* 36. https://www.nber.org/books-and-chapters/nber-macroeconomics-annual-2021-volume-36/climate-change-uncertainty-spillover-macroeconomy.

Barrage, Lint. 2020a. "Optimal Dynamic Carbon Taxes in a Climate–Economy Model with Distortionary Fiscal Policy." *Review of Economic Studies* 87, no. 1: 1–39. https://doi.org/10.1093/restud/rdz055.

Barrage, Lint. 2020b. "The Fiscal Costs of Climate Change." *American Economic Association Papers and Proceedings* 110 (May): 107–12. https://doi.org/10.1257/pandp.20201082.

Bauer, Michael, and Glenn Rudebusch. 2020. "Interest Rates under Falling Stars." *American Economic Review* 110, no. 5: 1316–54. https://doi.org/10.1257/aer.20171822.

Bauer, Michael, and Glenn Rudebusch. 2021. "The Rising Cost of Climate Change: Evidence from the Bond Market." *Review of Economics and Statistics*. https://direct.mit.edu/rest/article-abstract/doi/10.1162/rest_a_01109/107405/The-Rising-Cost-of-Climate-Change-Evidence-from?redirectedFrom=fulltext.

Borenstein, Severin, James Bushnell, Frank A. Wolak, and Matthew Zaragoza-Watkins. 2019. "Expecting the Unexpected: Emissions Uncertainty and Environmental Market Design." *American Economic Review* 109, no. 11: 3953–77. https://doi.org/10.1257/aer.20161218.

Bradford, David F. 1975. "Constraints on Government Investment Opportunities and the Choice of Discount Rate." *American Economic Review* 65, no. 5: 887–99.

Bressler, R. Daniel. 2021. "The Mortality Cost of Carbon." *Nature Communications* 12, no. 1: 4467. https://doi.org/10.1038/s41467-021-24487-w.

Caballero, Ricardo J., Emmanuel Farhi, and Pierre-Olivier Gourinchas. 2017. "Rents, Technical Change, and Risk Premia Accounting for Secular Trends in Interest Rates, Returns on Capital, Earning Yields, and Factor Shares." *American Economic Review* 107, no. 5: 614–20. https://doi.org/10.1257/aer.p20171036.

Cai, Yongyang, and Thomas S. Lontzek. 2019. "The Social Cost of Carbon with Economic and Climate Risks." *Journal of Political Economy* 127, no. 6: 2684–734. https://doi.org/10.1086/701890.

Calvin, Katherine, Bryan K. Mignone, Haroon S. Kheshgi, Abigail C. Snyder, Pralit Patel, Marshall Wise, Leon E. Clarke, and Jae Edmonds. 2020. "Global Market and Economic Welfare Implications of Changes in Agricultural Yields Due to Climate Change." *Climate Change Economics* 11, no. 1: 2050005. https://doi.org/10.1142/S2010007820500050.

Carleton, Tamma, Michael Delgado, Michael Greenstone, Trevor Houser, Solomon Hsiang, Andrew Hultgren, and others. 2018. "Valuing the Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits." Working Paper. Social Science Research Network, August 1. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3224365.

Carleton, Tamma, and Michael Greenstone. 2021. "Updating the United States Government's Social Cost of Carbon." Working Paper. Social Science Research Network, January 14. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3764255.

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* 101, no. 5: 933–49. https://doi.org/10.1162/rest_a_00821.

Christensen, Peter, Kenneth Gillingham, and William D. Nordhaus. 2018. "Uncertainty in Forecasts of Long-Run Economic Growth." *Proceedings of the National Academy of Sciences* 115, no. 21: 5409–14. https://doi.org/10.1073/pnas.1713628115.

Clarke, Leon, Jiyong Eom, Elke Hodson Marten, Russell Horowitz, Page Kyle, Robert Link, Bryan K. Mignone, Anupriya Mundra, and Yuyu Zhou. 2018. "Effects of Long-Term Climate Change on Global Building Energy Expenditures." *Energy Economics* 72 (May): 667–77. https://doi.org/10.1016/j.eneco.2018.01.003.

Clarke, Leon, and John Weyant. 2009. "Introduction to the EMF 22 Special Issue on Climate Change Control Scenarios." *Energy Economics* 31, suppl. 2: S63. https://doi.org/10.1016/j.eneco.2009.10.014.

Colson, Abigail R., and Roger M. Cooke. 2017. "Cross Validation for the Classical Model of Structured Expert Judgment." *Reliability Engineering and System Safety* 163 (July): 109–20. https://doi.org/10.1016/j.ress.2017.02.003.

Colson, Abigail R., and Roger M. Cooke. 2018. "Expert Elicitation: Using the Classical Model to Validate Experts' Judgments." *Review of Environmental Economics and Policy* 12, no. 1: 113–32. https://doi.org/10.1093/reep/rex022.

15985-04a_Rennert-4thPgs.indd 268 5/12/22 12:08 PM

Commoner, Barry. 1972. "The Environmental Cost of Economic Growth." In *Economics of the Environment*, edited by Robert Dorfman and Nancy S. Dorfman. New York: W. W. Norton.

Cooke, Roger M. 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. New York: Oxford University Press.

Cooke, Roger M. 2013. "Uncertainty Analysis Comes to Integrated Assessment Models for Climate Change . . . and Conversely." *Climatic Change* 117, no. 3: 467–79. https://doi.org/10.1007/s10584-012-0634-y.

Cooke, Roger M., Deniz Marti, and Thomas Mazzuchi. 2021. "Expert Forecasting with and without Uncertainty Quantification and Weighting: What Do the Data Say?" *International Journal of Forecasting* 37, no. 1: 378–87. https://doi.org/10.1016/j.ijforecast.2020.06.007.

Council of Economic Advisers (CEA). 2017. "Discounting for Public Policy: Theory and Recent Evidence on the Merits of Updating the Discount Rate." Issue Brief. https://obamawhitehouse.archives.gov/sites/default/files/page/files/201701_cea_discounting_issue_brief.pdf.

Csiszar, I. 1975. "Divergence Geometry of Probability Distributions and Minimization Problems." *Annals of Probability* 3, no. 1: 146–58.

Del Negro, Marco, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti. 2017. "Safety, Liquidity, and the Natural Rate of Interest." *Brookings Papers on Economic Activity*, Spring, 235–316. https://doi.org/10.1353/eca.2017.0003.

Diaz, Delavane B. 2016. "Estimating Global Damages from Sea Level Rise with the Coastal Impact and Adaptation Model (CIAM)." *Climatic Change* 137, nos. 1–2: 143–56. https://doi.org/10.1007/s10584-016-1675-4.

Diaz, Delavane B., and Frances Moore. 2017. "Quantifying the Economic Risks of Climate Change." *Nature Climate Change* 7, no. 11: 774–82. https://doi.org/10.1038/nclimate3411.

Dietz, Simon, Christian Gollier, and Louise Kessler. 2018. "The Climate Beta." *Journal of Environmental Economics and Management* 87 (January): 258–74. https://doi.org/10.1016/j.jeem.2017.07.005.

Dietz, Simon, James Rising, Thomas Stoerk, and Gernot Wagner. 2021. "Economic Impacts of Tipping Points in the Climate System." *Proceedings of the National Academy of Sciences* 118, no. 34. https://doi.org/10.1073/pnas.2103081118.

Drèze, Jacques H. 1974. "Discount Rates and Public Investment: A Post-Scriptum." *Economica* 41, no. 161: 52–61. https://doi.org/10.2307/2553422.

Drupp, Moritz A., Mark C. Freeman, Ben Groom, and Frikk Nesje. 2018. "Discounting Disentangled." *American Economic Journal: Economic Policy* 10, no. 4: 109–34. https://doi.org/10.1257/pol.20160240.

Economist. 2017. "The EPA Is Rewriting the Most Important Number in Climate Economics." November 16. https://www.economist.com/united-states/2017/11/16/the-epa-is-rewriting-the-most-important-number-in-climate-economics.

Errickson, Frank, Klaus Keller, William D. Collins, Vivek Srikrishnan, and David Anthoff. 2021. "Equity Is More Important for the Social Cost of Methane than Climate Uncertainty." *Nature* 592:564–70.

Fosdick, Bailey, and Adrian E. Raftery. 2014. "Regional Probabilistic Fertility Forecasting by Modeling Between-Country Correlations." *Demographic Research* 30, no. 35: 1011–34. https://doi.org/10.4054/DemRes.2014.30.35.

Gasparrini, Antonio, Yuming Guo, Francesco Sera, Ana Maria Vicedo-Cabrera, Veronika Huber, Shilu Tong, and others. 2017. "Projections of Temperature-Related Excess Mortality under Climate Change Scenarios." *Lancet Planetary Health* 1, no. 9: E360–67. https://doi.org/10.1016/S2542-5196(17)30156-0.

Gerland, Patrick, Adrian E. Raftery, Hana Ševčíková, Nan Li, Danan Gu, Thomas Spoorenberg, and others. 2014. "World Population Stabilization Unlikely This Century." *Science* 346, no. 6206: 234–37. https://doi.org/10.1126/science.1257469.

Gietel-Basten, Stuart, Wolfgang Lutz, and Sergei Scherbov. 2013. "Very Long Range Global Population Scenarios to 2300 and the Implications of Sustained Low Fertility." *Demographic Research* 28, no. 39: 1145–66. https://doi.org/10.4054/DemRes.2013.28.39.

Giglio, Stefano, Matteo Maggiori, Krishna Rao, Johannes Stroebel, and Andreas Weber. 2021. "Climate Change and Long-Run Discount Rates: Evidence from Real Estate." *Review of Financial Studies* 34, no. 8: 3527–71.

Giglio, Stefano, Matteo Maggiori, and Johannes Stroebel. 2015. "Very Long-Run Discount Rates." *Quarterly Journal of Economics* 130, no. 1: 1–53. https://doi.org/10.1093/qje/qju036.

Gollier, Christian. 2008. "Discounting with Fat-Tailed Economic Growth." *Journal of Risk and Uncertainty* 37, nos. 2–3: 171–86. https://doi.org/10.1007/s11166-008-9050-0.

Gollier, Christian. 2014. "Discounting and Growth." *American Economic Review* 104, no. 5: 534–37. https://doi.org/10.1257/aer.104.5.534.

Goulder, Lawrence H. 1995. "Environmental Taxation and the Double Dividend: A Reader's Guide." *International Tax and Public Finance* 2, no. 2: 157–83. https://doi.org/10.1007/BF00877495.

Harberger, Arnold C. 1972. "On Measuring the Social Opportunity Cost of Public Funds." In *Project Evaluation: Collected Papers*. London: Palgrave Macmillan.

Hausfather, Zeke, and Glen P. Peters. 2020. "RCP8.5 Is a Problematic Scenario for Near-Term Emissions." *Proceedings of the National Academy of Sciences* 117, no. 45: 27791–92. https://doi.org/10.1073/pnas.2017124117.

Ho, Emily, David V. Budescu, Valentina Bosetti, Detlef P. van Vuuren, and Klaus Keller. 2019. "Not All Carbon Dioxide Emission Scenarios Are Equally Likely: A Subjective Expert Assessment." *Climatic Change* 155, no. 4: 545–61. https://doi.org/10.1007/s10584-019-02500-y.

Interagency Working Group on the Social Cost of Carbon (IWG). 2010. "Social Cost of Carbon for Regulatory Impact Analysis under Executive Order 12866." Washington: Author. https://www.epa.gov/sites/default/files/2016-12/documents/scc_tsd_2010.pdf.

Interagency Working Group on the Social Cost of Greenhouse Gases (IWG). 2016. "Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis under Executive Order 12866." Washington: Author. https://19january

2017snapshot.epa.gov/sites/production/files/2016-12/documents/sc_co2_tsd_august_2016.pdf.

Interagency Working Group on the Social Cost of Greenhouse Gases (IWG). 2021. "Social Cost of Carbon, Methane, and Nitrous Oxide Interim Estimates under Executive Order 13990." Washington: Author. https://www.whitehouse.gov/wp-content/uploads/2021/02/TechnicalSupportDocument_SocialCostofCarbonMethaneNitrousOxide.pdf?source=email.

Johannsen, Benjamin K., and Elmar Mertens. 2016. "A Time Series Model of Interest Rates with the Effective Lower Bound." Finance and Economics Discussion Series. Washington: Board of Governors of the Federal Reserve System. https://doi.org/10.17016/FEDS.2016.033.

Johnson, Keith. 2009. "Waxman-Markey: Benefits Far Outweigh Costs, New Study Finds." *Wall Street Journal*, September 8. https://www.wsj.com/articles/BL-EB-7156.

Johnson, N. L. 1949. "Systems of Frequency Curves Generated by Methods of Translation." *Biometrika* 36, nos. 1–2: 149–76. https://doi.org/10.2307/2332539.

Kikstra, Jarmo, Paul Waidelich, James Rising, Dmitry Yumashev, Chris Hope, and Chris Brierley. 2021. "The Social Cost of Carbon Dioxide under Climate-Economy Feedbacks." *Environmental Research Letters* 16, no. 9: 094037. https://doi.org/10.13140/RG.2.2.34696.55044.

Kiley, Michael T. 2020. "What Can the Data Tell Us about the Equilibrium Real Interest Rate?" *International Journal of Central Banking* 16, no. 3: 181–209.

Kopp, Robert E., and Bryan K. Mignone. 2012. "The U.S. Government's Social Cost of Carbon Estimates after Their First Two Years: Pathways for Improvement." *Economics* 6, no. 1: 20120015. https://doi.org/10.5018/economics-ejournal.ja.2012-15.

Laubach, Thomas, and John C. Williams. 2016. "Measuring the Natural Rate of Interest Redux." *Business Economics* 51:57–67.

Leach, Nicholas J., Stuart Jenkins, Zebedee Nicholls, Christopher J. Smith, John Lynch, Michelle Cain, Tristram Walsh, Bill Wu, Junichi Tsutsui, and Myles R. Allen. 2021. "FaIRv2.0.0: A Generalized Impulse Response Model for Climate Uncertainty and Future Scenario Exploration." *Geoscientific Model Development* 14, no. 5: 3007–36. https://doi.org/10.5194/gmd-14-3007-2021.

Lemoine, Derek. 2021. "The Climate Risk Premium: How Uncertainty Affects the Social Cost of Carbon." *Journal of the Association of Environmental and Resource Economists* 8, no. 1: 27–57. https://doi.org/10.1086/710667.

Li, Qingran, and William A. Pizer. 2021. "Use of the Consumption Discount Rate for Public Policy over the Distant Future." *Journal of Environmental Economics and Management* 107 (May): 102428. https://doi.org/10.1016/j.jeem.2021.102428.

Liu, Peiran, and Adrian E. Raftery. 2020. "Accounting for Uncertainty about Past Values in Probabilistic Projections of the Total Fertility Rate for Most Countries." *Annals of Applied Statistics* 14, no. 2: 685–705. https://doi.org/10.1214/19-AOAS1294.

Liu, Peiran, and Adrian E. Raftery. 2021. "Country-Based Emissions Reductions Should Increase by 80% beyond Nationally Determined Contributions to Meet the 2°C Target." *Communications Earth and Environment* 2, article 29.

Marglin, Stephen A. 1963a. "The Opportunity Costs of Public Investment." *Quarterly Journal of Economics* 77, no. 2: 274–89. https://doi.org/10.2307/1884403.

Marglin, Stephen A. 1963b. "The Social Rate of Discount and the Optimal Rate of Investment." *Quarterly Journal of Economics* 77(1): 95–111. https://doi.org/10.2307/1879374.

Millar, Richard J., Zebedee R. Nicholls, Pierre Friedlingstein, and Myles R. Allen. 2017. "A Modified Impulse-Response Representation of the Global Near-Surface Air Temperature and Atmospheric Concentration Response to Carbon Dioxide Emissions." *Atmospheric Chemistry and Physics* 17, no. 11: 7213–28. https://doi.org/10.5194/acp-17-7213-2017.

Moore, Frances C., Uris Baldos, Thomas Hertel, and Delavane B. Diaz. 2017. "New Science of Climate Change Impacts on Agriculture Implies Higher Social Cost of Carbon." *Nature Communications* 8, article 1607. https://doi.org/10.1038/s41467-017-01792-x.

Müller, Ulrich K., James H. Stock, and Mark W. Watson. 2020. "An Econometric Model of International Growth Dynamics for Long-Horizon Forecasting." *Review of Economics and Statistics*. https://doi.org/10.1162/rest_a_00997.

Müller, Ulrich K., and Mark W. Watson. 2016. "Measuring Uncertainty about Long-Run Predictions." *Review of Economic Studies* 83, no. 4: 1711–40. https://doi.org/10.1093/restud/rdw003.

National Academies of Sciences, Engineering, and Medicine (NASEM). 2017. *Valuing Climate Damages: Updating Estimation of the Social Cost of Carbon Dioxide.* Washington: National Academies Press. https://doi.org/10.17226/24651.

Newell, Richard G., and William A. Pizer. 2003. "Discounting the Distant Future: How Much Do Uncertain Rates Increase Valuations?" *Journal of Environmental Economics and Management* 46, no. 1: 52–71.

Newell, Richard G., William A. Pizer, and Brian C. Prest. 2021. "A Discounting Rule for the Social Cost of Carbon." Working Paper. Washington: Resources for the Future. https://www.rff.org/publications/working-papers/a-discounting-rule-for-the-social-cost-of-carbon/.

Nicholls, Zebedee R. J., Malte Meinshausen, Jared Lewis, Robert Gieseke, Dietmar Dommenget, Kalyn Dorheim, and others. 2020. "Reduced Complexity Model Intercomparison Project Phase 1: Introduction and Evaluation of Global-Mean Temperature Response." *Geoscientific Model Development* 13, no. 11: 5175–90. https://doi.org/10.5194/gmd-13-5175-2020.

Nordhaus, William D. 1982. "How Fast Should We Graze the Global Commons?" *American Economic Review* 72, no. 2: 242–46.

Nordhaus, William D. 2017a. "Revisiting the Social Cost of Carbon." *Proceedings of the National Academy of Sciences* 114, no. 7: 1518–23. https://doi.org/10.1073/pnas.1609244114.

Nordhaus, William D. 2017b. "Evolution of Assessments of the Economics of Global Warming: Changes in the DICE Model, 1992–2017." Working Paper 23319.

Cambridge, Mass.: National Bureau of Economic Research. https://doi.org/10.3386/w23319.

Nordhaus, William D., and Paul Sztorc. 2013. *DICE 2013R: Introduction and User's Manual*, 2nd ed. http://www.econ.yale.edu/~nordhaus/homepage/homepage/documents/DICE_Manual_100413r1.pdf

Oeppen, Jim, and James W. Vaupel. 2002. "Demography: Broken Limits to Life Expectancy." *Science* 296, no. 5570: 1029–31. https://doi.org/10.1126/science.1069675.

Pigou, Arthur C. 1920. *The Economics of Welfare*. New York: Macmillan. https://www.palgrave.com/gp/book/9780230249318.

Pizer, William A. 2021. "A Shadow-Price-of-Capital Approach to Harmonize Discounting for Greenhouse Gases in Broader Benefit-Cost Analyses." Washington: Resources for the Future. https://media.rff.org/documents/OMB-2021-0006-0059_attachment_1.pdf.

Prest, Brian C. 2021. "Supply-Side Reforms to Oil and Gas Production on Federal Lands: Modeling the Implications for Climate Emissions, Revenues, and Production Shifts." Working Paper. Washington: Resources for the Future. https://www.rff.org/publications/working-papers/supply-side-reforms-oil-and-gas-production-federal-lands/.

Prest, Brian C., and James H. Stock. 2021. "Climate Royalty Surcharges." Working Paper. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28564.

Preston, Samuel, Patrick Heuveline, and Michel Guillot. 2001. *Demography: Measuring and Modeling Population Processes*. Malden, Mass.: Blackwell.

Rachel, Łukasz, and Lawrence H. Summers. 2019. "On Falling Neutral Real Rates, Fiscal Policy, and the Risk of Secular Stagnation." *Brookings Papers on Economic Activity*, Spring, 1–76.

Raftery, Adrian E. 2021. "The Dip in the US Birthrate Isn't a Crisis, but the Fall in Immigration May Be." *Conversation*, June 21. http://theconversation.com/the-dip-in-the-us-birthrate-isnt-a-crisis-but-the-fall-in-immigration-may-be-161169.

Raftery, Adrian E., Leontine Alkema, and Patrick Gerland. 2014. "Bayesian Population Projections for the United Nations." *Statistical Science* 29, no. 1: 58–68. https://doi.org/10.1214/13-STS419.

Raftery, Adrian E., Jennifer L. Chunn, Patrick Gerland, and Hana Ševčíková. 2013. "Bayesian Probabilistic Projections of Life Expectancy for All Countries." *Demography* 50, no. 3: 777–801. https://doi.org/10.1007/s13524-012-0193-x.

Raftery, Adrian E., Nevena Lalic, and Patrick Gerland. 2014. "Joint Probabilistic Projection of Female and Male Life Expectancy." *Demographic Research* 30, no. 27: 795–822. https://doi.org/10.4054/DemRes.2014.30.27.

Raftery, Adrian E., Nan Li, Hana Ševčíková, Patrick Gerland, and Gerhard K. Heilig. 2012. "Bayesian Probabilistic Population Projections for All Countries." *Proceedings of the National Academy of Sciences* 109, no. 35: 13915–21. https://doi.org/10.1073/pnas.1211452109.

Raftery, Adrian E., and Hana Ševčíková. 2021. "Probabilistic Population Forecasting: Short to Very Long-Term." *International Journal of Forecasting*. https://www.sciencedirect.com/science/article/pii/S0169207021001394?via%3Dihub.

Raftery, Adrian E., Alec Zimmer, Dargan M. W. Frierson, Richard Startz, and Peiran Liu. 2017. "Less than 2°C Warming by 2100 Unlikely." *Nature Climate Change* 7, no. 9: 637–41. https://doi.org/10.1038/nclimate3352.

Raimi, Daniel. 2021. "Effects of Climate Change on Heat- and Cold-Related Mortality: A Literature Review to Inform Updated Estimates of the Social Cost of Carbon." Working Paper. Washington: Resources for the Future. https://www.rff.org/publications/working-papers/effects-of-climate-change-on-heat-and-cold-related-mortality-a-literature-review-to-inform-updated-estimates-of-the-social-cost-of-carbon/.

Ramsey, F. P. 1928. "A Mathematical Theory of Saving." *Economic Journal* 38, no. 152: 543–59. https://doi.org/10.2307/2224098.

Resources for the Future (RFF) and New York State Energy Research and Development Authority (NYSERDA). 2021. *Estimating the Value of Carbon: Two Approaches*. Washington: Resources for the Future. https://www.rff.org/publications/reports/estimating-the-value-of-carbon-two-approaches/.

Riahi, Keywan, Detlef P. van Vuuren, Elmar Kriegler, Jae Edmonds, Brian C. O'Neill, Shinichiro Fujimori, and others. 2017. "The Shared Socioeconomic Pathways and Their Energy, Land Use, and Greenhouse Gas Emissions Implications: An Overview." *Global Environmental Change* 42 (January): 153–68. https://doi.org/10.1016/j.gloenvcha.2016.05.009.

Robinson, Lisa A., James K. Hammitt, and Lucy O'Keeffe. 2019. "Valuing Mortality Risk Reductions in Global Benefit-Cost Analysis." *Journal of Benefit-Cost Analysis* 10, no. S1: 15–50.

Rodgers, Michael, and Brandon Dubov. 2021. "US Tax Credit Encourages Investment in Carbon Capture and Storage." White and Case Insights, January 29. https://www.whitecase.com/publications/insight/carbon-capture/us-tax-credit-encourages-investment.

Rose, Steven K., Delavane B. Diaz, and Geoffrey J. Blanford. 2017. "Understanding the Social Cost of Carbon: A Model Diagnostic and Inter-Comparison Study." *Climate Change Economics* 8, no. 2: 1750009. https://doi.org/10.1142/S2010007817500099.

Rose, Steven K., Delavane Turner, Geoffrey J. Blanford, John Bistline, Francisco C. de la Chesnaye, and Tom Wilson. 2014. "Understanding the Social Cost of Carbon: A Technical Assessment." Washington: Electric Power Research Institute.

Roston, Eric. 2021. "The Most Important Number You've Never Heard Of." *Bloomberg*, January 22. https://www.bloomberg.com/news/articles/2021-01-22/how-do-you-put-a-price-on-climate-change-michael-greenstone-knows.

Sandmo, Agnar, and Jacques H. Drèze. 1971. "Discount Rates for Public Investment in Closed and Open Economies." *Economica* 38, no. 152: 395–412. https://doi.org/10.2307/2551880.

Sjaastad, Larry A., and Daniel L. Wisecarver. 1977. "The Social Cost of Public Finance." *Journal of Political Economy* 85, no. 3: 513–47. https://doi.org/10.1086/260582.

Stammer, Detlef, Anita Engels, Jochem Marotzke, Eduardo Gresse, Christopher Hedemann, and Jan Petzold. 2021. *Hamburg Climate Futures Outlook: Assessing the Plausibility of Deep Decarbonization by 2050.* Hamburg: Universität Hamburg. https://doi.org/10.25592/uhhfdm.9104.

Stern, Nicholas. 2007. *The Economics of Climate Change: The Stern Review*. Cambridge: Cambridge University Press. https://www.cambridge.org/us/academic/subjects/earth-and-environmental-science/climatology-and-climate-change/economics-climate-change-stern-review, https://www.cambridge.org/us/academic/subjects/earth-and-environmental-science/climatology-and-climate-change.

United Nations. 2004. *World Population to 2300*. New York: UN Department of Economic and Social Affairs. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/documents/2020/Jan/un_2002_world_population_to_2300.pdf.

United Nations. 2019. *World Population Prospects 2019*. New York: UN Department of Economic and Social Affairs. https://population.un.org/wpp/.

Vallin, J., and G. Caselli. 1997. "Towards a New Horizon in Demographic Trends: The Combined Effects of 150 Years Life Expectancy and New Fertility Models." In *Longevity: To the Limits and Beyond*, edited by J. M. Robine, J. W. Vaupel, B. Jeune, and M. Allard. Berlin: Springer-Verlag.

Weitzman, Martin L. 1998. "Why the Far-Distant Future Should Be Discounted at Its Lowest Possible Rate." *Journal of Environmental Economics and Management* 36, no. 3: 201–8. https://doi.org/10.1006/jeem.1998.1052.

Weitzman, Martin L. 2011. "Fat-Tailed Uncertainty in the Economics of Catastrophic Climate Change." *Review of Environmental Economics and Policy* 5, no. 2: 275–92. https://doi.org/10.1093/reep/rer006.

Weitzman, Martin L. 2014. "Fat Tails and the Social Cost of Carbon." *American Economic Review* 104, no. 5: 544–46. https://doi.org/10.1257/aer.104.5.544.

White House. 2003. Circular A-4. Washington: Office of Management and Budget. https://obamawhitehouse.archives.gov/omb/circulars_a004_a-4.

White House. 2021a. "Modernizing Regulatory Review." Brief. Washington: Author. https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/modernizing-regulatory-review/.

White House. 2021b. "Executive Order on Climate-Related Financial Risk." Brief. Washington: Author. https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/20/executive-order-on-climate-related-financial-risk/.

World Health Organization (WHO). 2014. *Quantitative Risk Assessment of the Effects of Climate Change on Selected Causes of Death, 2030s and 2050s*. Geneva: Author. https://apps.who.int/iris/bitstream/handle/10665/134014/9789241507691_eng.pdf?sequence=1&isAllowed=y.

Comments and Discussion

COMMENT BY

MICHAEL GREENSTONE For a century, large temperature increases have been observed and have had a significant impact on the overall climate (IPCC 2021). Greenhouse gas emissions, including largely CO_2, play a deterministic part in rising temperatures. While reduced greenhouse gas emissions can be costly, associated mitigation in rising temperatures can lead to significant reductions in climate damages and net improvements to welfare.

The social cost of carbon (SCC) is a critical input into assessing whether potential climate policies have benefits that exceed their costs. The SCC is the monetized value of all future net damages associated with the release of an additional ton of CO_2. The SCC, therefore, provides a measure of how much society should be willing to pay for a one-ton reduction in CO_2 emissions and allows policymakers to conduct a comparison of a regulation's benefits and costs, both measured in dollars. The SCC became a key part of US climate policy in 2010 (Greenstone, Kopits, and Wolverton 2013; IWG 2013) and has been used extensively in the United States and internationally since then. The basis for the US government's estimate of the SCC is derived from William Nordhaus's seminal research estimating the costs of climate damages and the SCC, as well as the Climate Framework for Uncertainty, Negotiation and Distribution (FUND) and Policy Analysis of the Greenhouse Effect (PAGE) integrated assessment models (Nordhaus 1992; Anthoff and Tol 2014; Hope 2011). To bastardize Winston Churchill's famous quote about democracy—as of 2010, the integrated assessment models were the worst approach to estimating climate damages, except for all the others that have been tried. (Churchill's quote is

15985-04b_Rennert_Com&GD-6thPgs.indd 276 5/12/22 12:08 PM

"Democracy is the worst form of Government except for all those other forms that have been tried from time to time.")[1]

It is my great pleasure to discuss this paper by Rennert and others that suggests a new approach for the US government to update the SCC. The authors' approach emphasizes socioeconomic uncertainty and its correlation with damages. My goal in this comment is to situate their contribution in the broader context of a holistic approach to updating the US government's SCC, underline drawbacks of the previous approach, and suggest criteria for the SCC calculation that makes it consistent with advances in the literature, economic theory, and policy objectives.

Overall, my conclusion is that the authors have taken an important step in fixing what ails the SCC, but their improvements need to be digested and examined by the scientific community. Further, to this point, their solutions fail to exploit the advances in damage estimation, which many believe to be the area where the most progress has been made in the last ten to fifteen years. Overall, this is an important contribution but more is needed to return the US government's SCC to the frontier of scientific understanding about climate damages.

BACKGROUND

The SCC in climate policy. In the United States, most major legislation requires agencies to conduct cost-benefit analyses. For policies aimed to reduce CO_2 emissions, such analysis heavily relies on the SCC. During the Obama administration, the SCC was set at $51 by the Interagency Working Group (IWG 2013). Setting out to roll back environmental regulations, the Trump administration lowered this number to $1–$8 by restricting damages to domestic ones and applying higher discount rates (Plumer 2018). Currently, the Biden administration has returned the SCC to $51 as an "interim" value and is actively working to update it.

The SCC extensively influences public policy. Through 2017, it had been used in analyzing the value of more than eighty regulations with gross benefits exceeding a trillion dollars (Nordhaus 2017). Moreover, at least eleven state governments, including Illinois and New York, use the SCC to value zero-emissions credits paid to clean energy producers (Rennert and

1. International Churchill Society, "Quotes," https://winstonchurchill.org/resources/quotes/the-worst-form-of-government/.

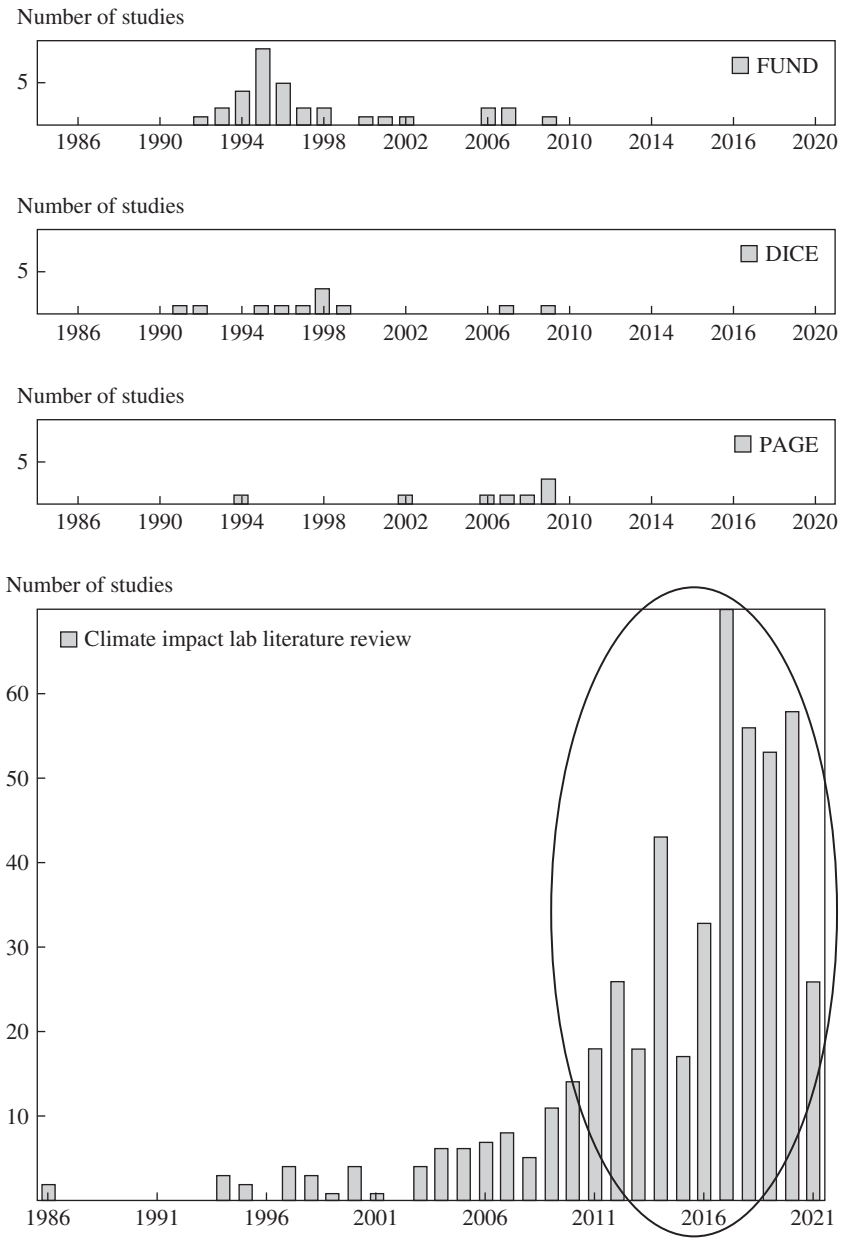
Kingdon 2019). The SCC had also been implemented internationally—countries like Canada, France, Germany, Mexico, Norway, and the United Kingdom all implemented the SCC to some extent (Institute for Policy Integrity 2014). Finally, meaningful US mitigation efforts facilitate international climate negotiations and lead to significant reductions in emissions from other countries (Houser and Larsen 2021).

The US government's SCC is no longer on the frontier of understanding. I co-led the IWG in 2009–2010 along with Cass Sunstein. Nordhaus (1992) was incredibly influential in shaping the economics profession's thinking and giving us a framework to think about the SCC and climate damages. This was achieved through the Dynamic Integrated Climate Change (DICE) model, which is generically referred to as an integrated assessment model along with PAGE and FUND. These three models all date back to the 1990s and formed the basis of the SCC. Even though all three models were somewhat dated by 2010, it was reasonable to conclude that the SCC reflected the frontier of understanding because the economics profession had not done much to update them since their creation.

In the intervening dozen years, these three integrated assessment models, and the resulting SCC, have fallen behind this frontier in several key areas. First, the set of models is highly reliant on expert judgment; there is a limited set of people with knowledge of what goes on inside the models, making replication, validation, and improvements challenging. This kind of small-club approach is not the best way to make scientific progress. Second, when these three models were created, computational resources were relatively limited, which forced researchers to make several simplifying assumptions. For example, the estimated relationship between human well-being and changes in temperature in these models is based on quite limited data and a heavy reliance on functional form assumptions. In this respect, they generally have not taken advantage of the robust and rapidly growing climate damages literature (Deschênes and Greenstone 2011) that has emerged in the past two decades (see figure 1). Similarly, the underlying climate models are dated and fail to capture many climate dynamics. Third, integrated assessment models used deterministic models in most cases, and hence such models incompletely accounted for uncertainty in their estimates. Finally, these models produced highly aggregated estimates of climate impacts. Even the most disaggregated model, FUND, has only sixteen regions, assuming, for instance, that climate change will affect Miami and Minneapolis identically. Recent research has established that the impacts of climate change are highly heterogeneous, and this heterogeneity matters for the SCC calculation. For instance, Hsiang and others

15985-04b_Rennert_Com&GD-6thPgs.indd 278 5/12/22 12:08 PM

Figure 1. Empirical Publications Informing the Previous Generation of Models



Source: Climate Impact Lab.

Note: Foundational research for models used for the interim SCC dates to the 1990s and hence misses out on the growing body of literature on damage function estimation, for example (Carleton and others 2020).

(2017) demonstrate that the projected end-of-century climate damages are nine times greater in the poorest 5 percent of US counties than in the richest 5 percent. Consistent with these findings, Carleton and others (2020) find significant differences in the projected change in mortality risk both across and within countries.

Given the amount of time that has passed, it is not surprising that the SCC is due for an update. Indeed, the original IWG suggested that the SCC should be updated regularly to reflect advances in the understanding of key components of the calculation (IWG 2010). The need for the update was pointed out again seven years later by the National Academies of Sciences, Engineering, and Medicine (NASEM 2017). Further, Carleton and Greenstone (2021) lay out a detailed plan for updating it. Finally, the SCC's legal durability relies on the estimation that is based on frontier science and economics.

UPDATING THE US GOVERNMENT'S APPROACH TO SCC ESTIMATION NASEM (2017) and Carleton and Greenstone (2021) explain that there are four key modules or ingredients in constructing the SCC. This section describes the authors' efforts in the paper in each of these areas, providing some context for their contributions.

Socioeconomic and emissions pathways. The current (as of March 2022) and past SCC calculations, which were developed using Energy Modeling Forum (EMF 22) scenarios (Clarke and others 2009), do not reflect the last decade of work in probabilistic scenario development. A specific concern has been that the SCC relied on five socioeconomic scenarios, which were equally weighted, and that they did not span the full uncertainty about economic growth, increases in greenhouse gas emissions, and population growth (NASEM 2017). Each of these socioeconomic variables is a key input into the SCC, meaning that it may not reflect the full range of expected variation in these variables.

For all three variables, the authors combined statistical projections of these variables with expert elicitation to generate probabilistic projections through 2300. These probabilistic projections are referred to as the RFF Socioeconomic Projections (RFF-SPs) and allow for substantially more socioeconomic uncertainty than was being captured previously.

In the case of economic growth, the authors rely on Müller, Stock, and Watson's (2019) statistical model of economic growth extended out to 2300. The model is derived from a data set that covers 113 countries from 1900 to 2017. Müller, Stock, and Watson's (2019) projections were coupled with formal expert elicitation about the "frontier of economic growth." The ten experts—Daron Acemoglu, Erik Brynjolfsson, Jean Chateau, Robert

15985-04b_Rennert_Com&GD-6thPgs.indd 280 5/12/22 12:08 PM

Gordon, Lant Pritchett, Melissa Dell, Mun Ho, Chad Jones, Dominique van der Mensbrugghe, and Pietro Peretto—were interviewed separately for roughly two hours each. It is worth noting that the authors "omit some projections in the extreme tails of Müller, Stock, and Watson's (2019) distribution that are outside the range of historical experience" and outside the range specified by the experts. This choice is motivated by the fact—as the authors notice—that "such low or high sustained growth rates would lead to global GDP/capita either falling by more than 90% between 2021 and 2300 (e.g., 0.99^{279}) or rising by a factor of more than 800,000 (1.05^{279}) implying a global average income of more than $10 billion per person" (online appendix).

With respect to population, the authors replaced EMF 22 projections with a probabilistic UN statistical model extended to 2300. Like with economic growth, due to expert disagreement concerning the projected lower bound on the total fertility rate, the model was further altered to account for population growth experts' views. Finally, to incorporate the uncertainty of emissions, greenhouse gas emissions projections from ten experts were paired with economic growth scenarios. In the case of population, nine experts were surveyed, while ten were surveyed for emissions trends.

Climate model. The integrated assessment models used in the SCC calculation represented economists' interpretation of climate change and did not reflect the last decade of modeling. The primary input into each model is equilibrium climate sensitivity, which represents the total warming realized from doubling CO_2 concentrations in the atmosphere. While equilibrium climate sensitivity has a tremendous impact on the interim SCC calculation, its actual value is not known with scientific precision.

Accounting for the best science available at the time, the IWG combined the equilibrium climate sensitivity estimates across all models by employing a probability distribution reflecting the likelihood of different possible climate outcomes at the end of the century adapted from the Intergovernmental Panel on Climate Change's fourth assessment report (IPCC 2007). Even so, integrated assessment models fail to precisely measure multiple links in the causal chain from CO_2 emissions to temperature change (Dietz and others 2021; Hänsel and others 2020; Montamat and Stock 2020; NASEM 2017). Particularly, these models significantly understated the speed of warming (Montamat and Stock 2020). For instance, increased carbon concentrations lead to warmer and more acidic oceans, which in turn makes them less effective at removing CO_2 from the atmosphere. The resulting positive feedback loop is missing from both the DICE and PAGE models (Dietz and others 2021). Furthermore, delayed warming

projected in these models likely results in a downward biased estimate of the SCC as warming further into the future is discounted more heavily.

The authors address this problem by incorporating the Finite Amplitude Impulse Response (FaIR) climate model (Millar and others 2017). This choice is consistent with NASEM's key climate module criteria for the SCC calculation (NASEM 2017). The model chosen by the authors generates climate projections consistent with comprehensive, frontier science models, such as the set of models composing the CMIP6 ensemble (Eyring and others 2016), and can be used to quantify uncertainty surrounding the impact of an additional metric ton of CO_2 on global mean surface temperature. Moreover, the FaIR model is computationally feasible, transparently documented, and is commonly used in the SCC updates (Carleton and others 2020; Dietz and others 2021; Hänsel and others 2020; Rode, Baker and others 2021; Rode, Carleton and others 2021). FaIR's main limitation is that it does not capture changes in global mean sea level rise. One promising, but imperfect, way to overcome this limitation is to use semiempirical models that enable the inclusion of damages due to projected sea level changes (Kopp and others 2016).

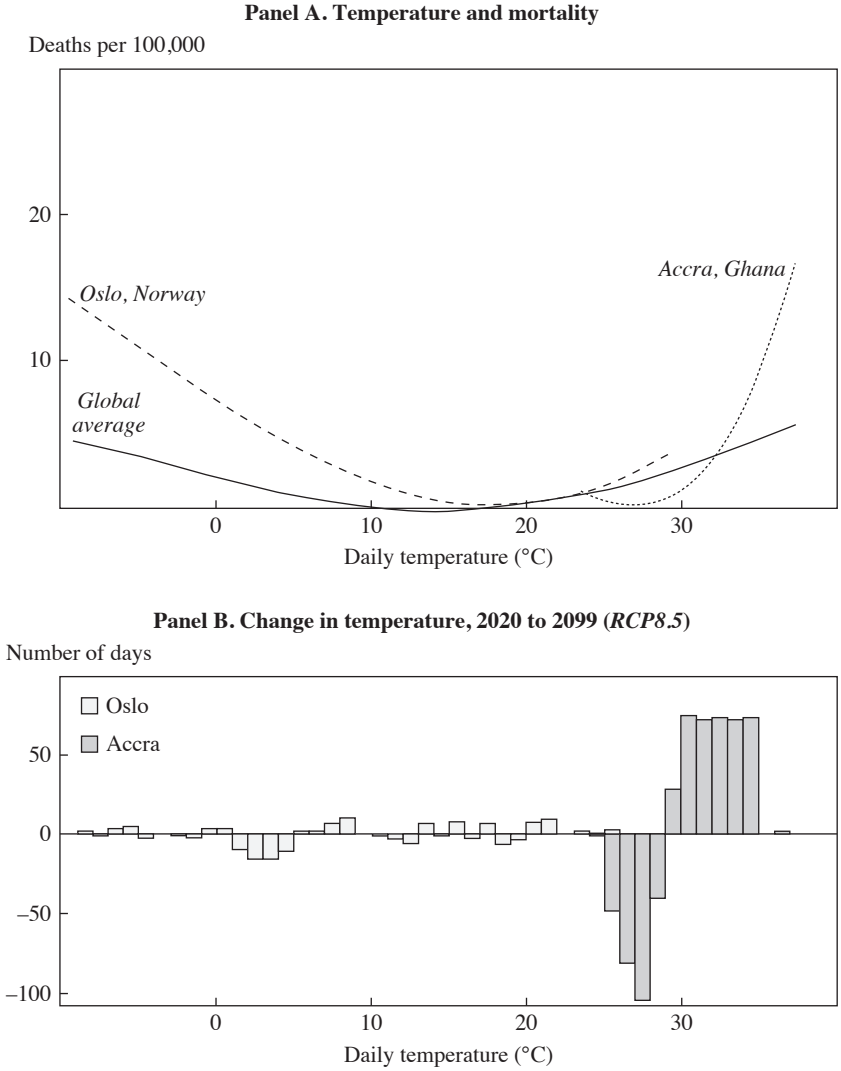
Damage functions. The next step in calculating the SCC is to make changes in the physical climate (e.g., temperature) and determine their impact on net economic damages. The relationship between economic damages and temperature change is known as a damage function. The previous generation of damage functions, which includes the FUND, DICE, and PAGE models, was developed in the 1990s and hence omits a rapidly growing literature (see figure 1). Indeed, my judgment is that this is the area with the greatest advances in understanding in the last few decades.

There are several shortcomings in the damage function used for the SCC calculation. First, the older models rely heavily on data from wealthy countries with temperate climates. This means that these models had to rely on ad hoc assumptions to create global damage functions because there simply was no support in the available data for the hot, poor, and hot and poor places where much of the world's population lives. Given the tremendous progress in data availability over the last few decades, there is no need to rely on ad hoc assumptions any longer; instead, there are now opportunities to rely on large-scale and globally representative data.

Second, there is substantial heterogeneity around the planet—what happens in Accra when hot temperatures arrive is vastly different from the effect of these temperatures in Oslo, for example (figure 2). This heterogeneity can in part be explained by nonlinear relationships between temperature, mortality, and adaptation. The DICE model currently used by the authors

15985-04b_Rennert_Com&GD-6thPgs.indd 282 5/12/22 12:08 PM

Figure 2. Climate Change Consequences Are Heterogeneous



Source: Adapted with permission from Carleton and Greenstone (2021), also available at SSRN: https://ssrn.com/abstract=3764255.

Note: The figure shows estimated mortality-temperature relationships for age 65 and older (panel A), as well as projected changes in temperature distribution, for Oslo, Norway, and Accra, Ghana (panel B).

as a placeholder ignores distributional impact by dividing the planet into no more than sixteen regions. To capture local nonlinearities, updated damage functions should be more granular, for example, Climate Impact Lab employs distributed computing using 24,378 regions.[2]

Third, even within a given region, economic and climate uncertainty is substantial in every aspect of damage function estimation: mortality, coastal, labor, agriculture, electricity, and other fuels, as can be seen in figure 3. Thus, updated damage function should account for heterogeneous effects of temperature across sectors, as well as econometric uncertainty.

Damage functions are the engine that drives the determination of the SCC, and the authors' reliance on older DICE damage functions means that their approach is behind the frontier. This is the case in three specific ways. First, it is now possible to rely on damage functions that are empirically founded and represent plausibly causal impacts of climate change on socioeconomic outcomes. Second, recent work has demonstrated that data representative of the global population, not just rich or temperate regions, are now available and can be used to estimate damage functions. Finally, damage functions should account for both estimated benefits and costs of future adaptive investments, and this is a hallmark of the Climate Impact Lab's estimation of damage functions represented in figure 3. See Carleton and Greenstone (2021) for a fuller discussion of these issues.

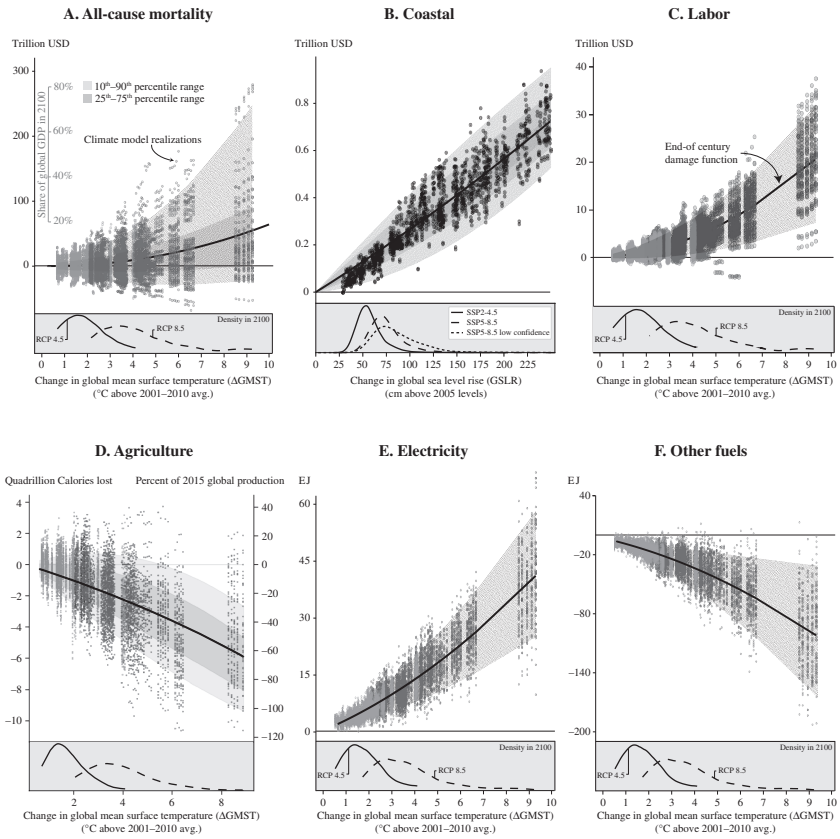
Discounting. CO_2 added to the atmosphere causes a stream of damages and benefits associated with a given trajectory of warming spanned over centuries. The choice of a discount rate is therefore highly consequential for determining the SCC. To date, the SCC relied on a central constant discount rate of 3 percent, following US government's guidance on the conduct of cost-benefit analysis.

This approach fails to account for several features of current economic thinking about discounting that are especially important in the climate context where greenhouse gases can influence the climate for centuries after their release. These features include: (1) the 3 percent figure is intended to reflect the riskless rate but that rate is now likely 2 percent or lower (Bauer and Rudebusch 2020, 2021); (2) uncertainty in the riskless discount rate would lead the discount rate to decline with the time horizon (Weitzman 1998), which constant discount rates do not capture; and (3) payoffs to emissions mitigation could be correlated with future income realizations,

2. As a disclosure, I am a codirector of Climate Impact Lab, which is a team of more than thirty researchers who aim to quantify the real-world costs of climate change. For details, visit https://impactlab.org/.

15985-04b_Rennert_Com&GD-6thPgs.indd 284 5/12/22 12:08 PM

Figure 3. Economic and Climate Uncertainty in Examined Sectors



Sources: Panel A adapted with permission from Carleton and others (2020); panel B adapted with permission from Depsky and others (forthcoming); panel C adapted with permission from Rode, Baker, and others (2021); panel D adapted with permission from Hultgren and others (forthcoming); panels E and F from Rode, Carleton, and others (2021), adapted by permission from Springer Nature.

Note: Black dots represent projected possible damages by the end of the century under the RCP 4.5 scenario (solid line represents the density of change in global mean surface temperature). Similarly, gray dots represent projected possible damages by the end of the century under the RCP 8.5 scenario (dashed line represents the density of change in global mean surface temperature). The change in temperature is in degrees Celsius above 2001–2010 average. The horizontal axis for coastal damages represents change in sea level in centimeters relative to 2005.

in which case there is effectively a "climate beta" and riskless rates are inappropriate (Gollier and Hammitt 2014).

The Ramsey (1928) equation provides a standard way to think about the intertemporal problem of discounting that can accommodate these limitations of constant discounting. It is:

$$r_t = \rho + \eta g_t,$$

where r_t represents the discount rate at time t, ρ is the pure rate of time preference, η is the coefficient of relative risk aversions, and g_t is the per capita growth in consumption at time t. When the growth rate is uncertain, as is the case with the RFF-SP probabilistic growth scenarios, then the average discount rate in year t, r_t, is also uncertain. Following Weitzman (1998), the present value of damages from an additional ton of emissions, MD_t, is then given by:

$$PV(MD_t) = E[e^{-(\rho+\eta g_t)t} MD_t].$$

This means that there is a stochastic discount factor (due to g_t) and that produces a declining certainty-equivalent risk-free rate.

An appealing feature of this approach is that it incorporates the climate beta because when $\eta > 0$, the discount factor is smallest when growth is largest and largest when growth is smallest. This captures the idea that a dollar of damages is more meaningful when we are relatively poor. To apply these insights and connect them to the current riskless rates, the authors follow Newell, Pizer, and Prest (2021) and choose values of (ρ, η) that are disciplined by the current riskless rate. In so doing, they ignore the available evidence of the values of these parameters and instead use observed interest rates to govern the choice of the parameters. This creates an inconsistency between their approach and the large body of literature that has, for example, estimated values for η that range from 1 to 4 but are generally centered around 2 (Gollier and Hammitt 2014). So this aspect of their approach has practical appeal, but it is not built on an especially solid foundation of evidence.

The authors apply this approach and demonstrate its value. An especially important finding is that when uncertainty in economic growth is incorporated (as the RFF-SPs do), then constant discounting produces values of the SCC that appear inappropriately high. This is because it places a relatively greater weight on damages that occur in good times, which does not fit the widespread evidence on the declining marginal utility of consumption. Put another way, it ignores the climate beta and leads to an upward bias in the SCC.

COMMENTS AND CONCLUSIONS The SCC has been overdue for a revision probably for almost a decade but certainly for at least five years since the 2017 NASEM report was issued. In many respects, this paper is a

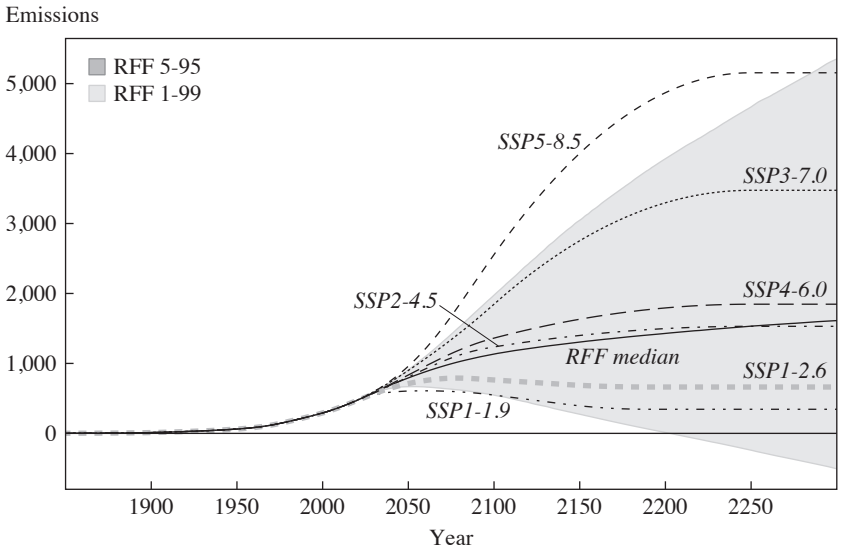
response to the near-term items that NASEM outlined. It updates the climate model, characterizes the uncertainty in projections about economic growth, population, and emissions, and implements a Ramsey-style approach to discounting that takes advantage of the characterization of uncertainty in the socioeconomics and that climate damages may be correlated with the overall economy. These are important accomplishments.

The straw that stirs this drink and this work's primary contribution are the socioeconomic projections. As a reminder, these boil down to projections of how these variables will evolve for the next three hundred years. This is a terribly difficult task but nevertheless a critical one for getting climate economics and policy right. The choice for developing multi-century estimates of how growth, population, and emissions will evolve essentially boils down to relying on expert judgment in one form or another (e.g., the probabilistic scenarios or the deterministic shared socioeconomic pathways), using statistical models to make projections, or some combination as the authors do in this paper.

I will confess to skepticism about the value of relying on responses from prominent researchers to a two-hour survey that in many instances does not relate to the core of their scientific work. The penalty of being wrong is essentially zero and internal consistency in answers across questions is not assured. Further, I think there is little disciplining the replies besides personal opinions, prejudices, and incomplete recollections of statistical models. And yet, the academic reputation of the respondents provides credibility to the entire exercise—credibility that I think is unwarranted given the challenges I have outlined here. In contrast, good statistical projections are disciplined in transparent ways. We may argue about the statistical approach, but it is at least clear what it was.

In this vein, the following figures plot statistical features of the distribution of future global CO_2 emissions from 10,000 joint population-GDP-emissions trajectories up to 2300 simulated in the paper. These trajectories are derived by sampling from the Resources for the Future distributions of future population, GDP, and emissions, which are constructed using a combination of prior studies and expert elicitation. In particular, emissions are paired one-to-one with each of 10,000 population-GDP trajectories based on a distribution constructed through a survey of experts. I note that separate distributions were specified for direct emissions and CO_2 removal through negative emissions technologies and that the authors generated net emissions by independently sampling from these distributions and summing. These figures are designed to show the basic statistical properties of the authors' projections, and they highlight some surprising features that at least partially arise due to the use of expert elicitation to shape probability distributions.

Figure 4. Cumulative Emissions through 2300, Alternative Scenarios



Source: Author's calculations.

Note: For each year indicated on the horizontal axis, the solid line represents the median value of cumulative emissions across 10,000 simulations given in the paper; dark gray and light gray shaded areas respectively show the 5th–95th and 1st–99th percentile ranges across these simulations. For comparison, dashed lines represent cumulative emissions trajectories under the deterministic scenarios used in the IPCC's Sixth Assessment Report (IPCC 2021).

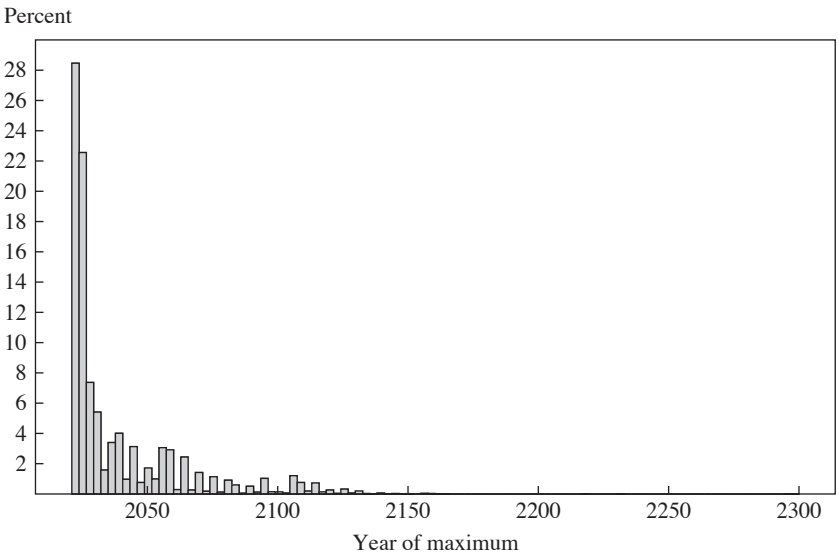
In figure 4, the solid line is the median across these 10,000 simulations, dark gray shading shows the 5th–95th percentile range, and light gray shading shows the 1st–99th percentile range. For comparison, the plots also show projections from the RCP/SSPs (dashed lines), which are the deterministic scenarios of socioeconomics and emissions used in the IPCC's Sixth Assessment Report (IPCC 2021).[3]

There are several noteworthy features of this figure. It is certainly interesting, and perhaps reassuring, that the median falls in the middle of the SSP-RCP combinations. However, I was especially struck by some of the patterns in the tail. For example, 4 percent of the projections have cumulative emissions that are negative by 2300. This can only be the case if

3. The five shared socioeconomic pathways (SSPs) (Riahi and others 2017) each contain a narrative of future conditions along with associated projections of future socioeconomic variables. In the IPCC's Sixth Assessment Report (IPCC 2021), these SSPs are paired with emissions pathways—representative concentration pathways (RCPs)—to generate combined socioeconomic and emissions scenarios.

15985-04b_Rennert_Com&GD-6thPgs.indd 288 5/12/22 12:08 PM

Figure 5. Histogram of Year of Peak Annual Emissions



Source: Author's calculations.

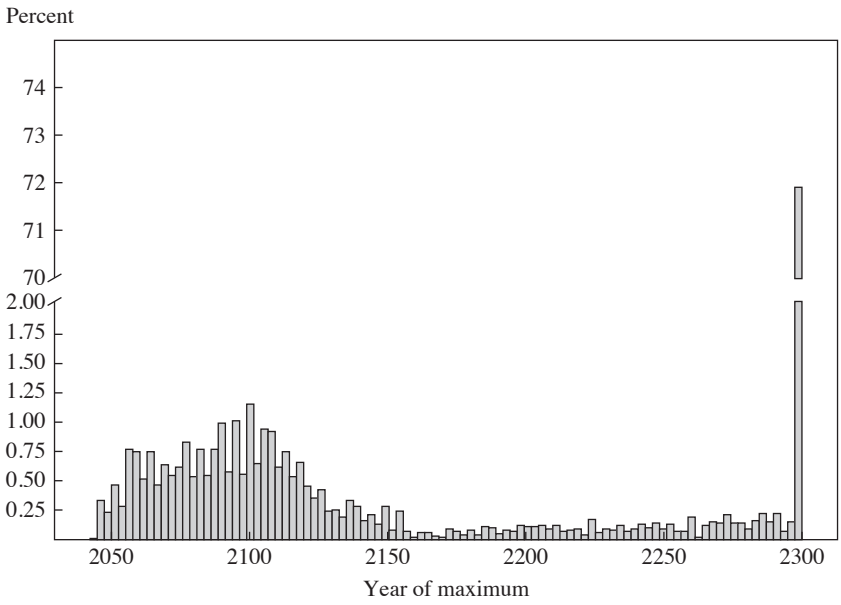
Note: The histogram represents the distribution of the year of peak annual emissions across the 10,000 simulations given in the paper. The height of bars indicates the percent of simulations that reach their peak annual emissions within a given year span.

economically and technically feasible carbon dioxide removal technologies arrive and are used so extensively that more CO_2 is removed from the atmosphere than was ever emitted since 1850. In effect, these projections assume that the optimal global temperature must be colder than preindustrial temperatures and societies continue to fund the operation of carbon dioxide removal machines until this colder optimum is reached. A not quite as astounding, but still surprising, feature of the projections is that 11.4 percent of them have cumulative emissions that are lower than current cumulative emissions. This too can only be explained by a massive use of carbon dioxide removal technologies. As a point of comparison, there are approximately zero technically and economically scalable examples of these technologies currently.

Figures 5 and 6 report histograms of the year that annual CO_2 emissions peak and the year that cumulative emissions peak (i.e., the sum of all future emissions is negative), which is also the year net-zero CO_2 emissions are achieved, respectively. In figure 5, annual global emissions peak by 2035 in 68 percent of the projections, by 2050 in 79 percent, and by 2100 in 95.7 percent. In figure 6, less than 1 percent of the projections have peak

15985-04b_Rennert_Com&GD-6thPgs.indd 289 5/12/22 12:08 PM

Figure 6. Histogram: Year when Net Zero Emissions Attained, RFF CO_2



Source: Author's calculations.

Note: The histogram represents the distribution of the year of peak cumulative emissions across the 10,000 simulations given in the paper. The height of bars indicates the percent of simulations that reach their peak cumulative emissions within a given year span.

cumulative emissions occur before 2050. By 2100, 13.7 percent of the projections have reached peak cumulative emissions. Without revealing my own expert judgments, I will note that this is at great odds with the Paris Climate Accords, which set a target of achieving net-zero by the middle of the twenty-first century. It is especially striking that 71.8 percent of the projections have not reached peak cumulative emissions by 2300, which would put any of the frequently discussed temperature change targets (e.g., 1.5 or 2.0 degrees Celsius) far out of reach. Personally, I am not quite sure what to make of these findings, but it would be instructive to have them interrogated by the academic community and to be explicit about the roles of the underlying statistical models and the expert judgment in producing them.

Overall, several of the findings from these figures surprised me. Does that mean that they are wrong? No. However, I think there is a strong case for opening these projections up to the research community so that they can be analyzed carefully. It would be especially interesting to compare them to projections that are based entirely on statistical models. Regardless of what

15985-04b_Rennert_Com&GD-6thPgs.indd 290 5/12/22 12:08 PM

conclusions are reached, the seriousness of the climate problem demands that such peer review be a part of the process of inserting expert elicitation–based projections into policy-relevant models of climate change.

Ultimately, my judgment is that all approaches to developing long-run socioeconomic projections are going to be unsatisfactory, but they are nevertheless necessary for devising a socially desirable policy to confront the climate change problem. The authors have made a careful effort to develop these projections. They should be carefully scrutinized with an eye toward how much weight, if any, to place on expert judgment. Regardless of what is chosen, the paper also deserves credit for demonstrating how to integrate uncertainty about these projections and recognizing that climate damages may be correlated with the overall economy into discounting through Ramsey-style discounting. Both of these contributions align with key near-term recommendations from the 2017 NASEM report.

I will close by noting that NASEM's 2017 medium- or long-run recommendations extended beyond the contributions in this paper and involved the incorporation of empirically founded damage functions into the calculation of the SCC. This NASEM recommendation came from the rapid advances in estimation techniques, data access, and computing that have made it possible to ground damage function estimation in data, rather than assumptions.

It is now possible to achieve these long-run NASEM goals. Trevor Houser, Solomon Hsiang, Robert Kopp, and I cofounded the Climate Impact Lab in 2014 to build climate damage functions empirically and use them to calculate the SCC. The guiding principles were a ruthless belief that the SCC should be based on the best available econometric evidence and that it must account for adaptation costs and benefits, be globally representative, rely on the best available climate models, and value uncertainty and unequal impacts. The result of this work is the development of the Data-driven Spatial Climate Impact Model (DSCIM), which is a modular system for computing the SCC and the global impacts of climate change at the level of 25,000 regions (e.g., a US county) around the world using data. DSCIM is built to be very flexible—for example, it can incorporate characterizations of econometric, climate, and socioeconomic uncertainty (including the authors' projections), value this uncertainty, implement essentially any approach to discounting, including the one the authors outline, and deliver estimates of climate damages where people live rather than at the global or country level.

What is ahead for SCC research? The to-do list is long but it certainly includes building out damage functions for more sectors (e.g., damages from

altered labor productivity, alterations in ecosystem services, migration, etc.), improving understanding of the costs and benefits of adaptation, understanding the interaction of impacts in sectors (e.g., agriculture and migration), and so much more. This is an exciting area of research with enormous implications for policy.

ACKNOWLEDGMENTS I thank Bogdan Mukhametkaliev and Jared Stolove for excellent research assistance. I thank the terrific Climate Impact Lab team with an especial acknowledgment of Tamma Carleton, Trevor Houser, Solomon Hsiang, Robert Kopp, Kelly McCusker, and Ashwin Rode for insightful comments on this paper.

REFERENCES FOR THE GREENSTONE COMMENT

Anthoff, David, and Richard S. J. Tol. 2014. "The Income Elasticity of the Impact of Climate Change." In *Is the Environment a Luxury?*, edited by Silvia Tiezzi and Chiara Martini. London: Routledge.

Bauer, Michael D., and Glenn D. Rudebusch. 2020. "Interest Rates under Falling Stars." *American Economic Review* 110, no. 5: 1316–54.

Bauer, Michael D., and Glenn D. Rudebusch. 2021. "The Rising Cost of Climate Change: Evidence from the Bond Market." *Review of Economics and Statistics*. https://direct.mit.edu/rest/article-abstract/doi/10.1162/rest_a_01109/107405/The-Rising-Cost-of-Climate-Change-Evidence-from?redirectedFrom=fulltext.

Carleton, Tamma A., and Michael Greenstone. 2021. "Updating the United States Government's Social Cost of Carbon." Working Paper. Social Science Research Network, January 14. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3764255.

Carleton, Tamma A., Amir Jina, Michael T. Delgado, Michael Greenstone, Trevor Houser, Solomon M. Hsiang, and others. 2020. "Valuing the Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits." Working Paper 27599. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27599.

Clarke, Leon, Jae Edmonds, Volker Krey, Richard Richels, Steven Rose, and Massimo Tavoni. 2009. "International Climate Policy Architectures: Overview of the EMF 22 International Scenarios." *Energy Economics* 31, no. 5: S64–81.

Depsky, Nicholas, Ian Bolliger, Daniel Allen, Jun Ho Choi, Michael Delgado, Michael Greenstone, Ali Hamidi, Trevor Houser, Robert E. Kopp, and Solomon Hsiang. Forthcoming. "An Open-Source Modeling Platform for Assessing Global Impacts of 21st Century Sea Level Rise." Working Paper.

Deschênes, Olivier, and Michael Greenstone. 2011. "Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US." *American Economic Journal: Applied Economics* 3, no. 4: 152–85.

Dietz, Simon, Frederick van der Ploeg, Armon Rezai, and Frank Venmans. 2021. "Are Economists Getting Climate Dynamics Right and Does It Matter?"

Journal of the Association of Environmental and Resource Economists 8, no. 5: 895–921.

Eyring, Veronika, Sandrine Bony, Gerald A. Meehl, Catherine A. Senior, Bjorn Stevens, Ronald J. Stouffer, and Karl E. Taylor. 2016. "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organization." *Geoscientific Model Development* 9, no. 5: 1937–58. doi:10.5194/gmd-9-1937-2016.

Gollier, Christian, and James K. Hammitt. 2014. "The Long-Run Discount Rate Controversy." *Annual Review of Resource Economics* 6, no. 1: 273–95.

Greenstone, Michael, Elizabeth Kopits, and Ann Wolverton. 2013. "Developing a Social Cost of Carbon for US Regulatory Analysis: A Methodology and Interpretation." *Review of Environmental Economics and Policy* 7, no. 1: 23–46.

Hänsel, Martin C., Moritz A. Drupp, Daniel J. A. Johansson, Frikk Nesje, Christian Azar, Mark C. Freeman, Ben Groom, and Thomas Sterner. 2020. "Climate Economics Support for the UN Climate Targets." *Nature Climate Change* 10:781–89.

Hope, Chris. 2011. "The PAGE09 Integrated Assessment Model: A Technical Description." Working Paper. Cambridge: University of Cambridge Judge Business School. https://www.jbs.cam.ac.uk/wp-content/uploads/2020/08/wp1104.pdf.

Houser, Trevor, and Kate Larsen. 2021. "Calculating the Climate Reciprocity Ratio for the US." Rhodium Group, January 21. https://rhg.com/research/climate-reciprocity-ratio/.

Hsiang, Solomon, Robert Kopp, Amir Jina, James Rising, Michael Delgado, Shashank Mohan, and others. 2017. "Estimating Economic Damage from Climate Change in the United States." *Science* 356, no. 6345: 1362–69.

Hultgren, Andrew, Tamma A. Carleton, Michael T. Delgado, Diana R. Gergel, Michael Greenstone, Trevor Houser, and others. Forthcoming. "Estimating Global Impacts to Agriculture from Climate Change Accounting for Adaptation." Working Paper.

Institute for Policy Integrity. 2014. "Social Cost of Carbon Pollution Fact Sheet." New York: New York University School of Law.

Interagency Working Group on the Social Cost of Carbon (IWG). 2010. "Social Cost of Carbon for Regulatory Impact Analysis under Executive Order 12866." Washington: Author. https://www.epa.gov/sites/default/files/2016-12/documents/scc_tsd_2010.pdf.

Interagency Working Group on the Social Cost of Carbon (IWG). 2013. "Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis under Executive Order 12866." Washington: Author.

Intergovernmental Panel on Climate Change (IPCC). 2007. *AR4 Climate Change 2007: Synthesis Report*. https://www.ipcc.ch/assessment-report/ar4/.

Intergovernmental Panel on Climate Change (IPCC). 2021. *Climate Change 2021: The Physical Science Basis*. Cambridge: Cambridge University Press, in press.

Kopp, Robert E., Andrew C. Kemp, Klaus Bittermann, Benjamin P. Horton, Jeffrey P. Donnelly, W. Roland Gehrels, Carling C. Hay, Jerry X. Mitrovica, Eric D. Morrow, and Stefan Rahmstorf. 2016. "Temperature-Driven Global Sea-Level Variability in the Common Era." *Proceedings of the National Academy of Sciences* 113, no. 11. https://doi.org/10.1073/pnas.1517056113.

Millar, Richard J., Zebedee R. Nicholls, Pierre Friedlingstein, and Myles R. Allen. 2017. "A Modified Impulse-Response Representation of the Global Near-Surface Air Temperature and Atmospheric Concentration Response to Carbon Dioxide Emissions." *Atmospheric Chemistry and Physics* 17, no. 11: 7213–28. https://doi.org/10.5194/acp-17-7213-2017.

Montamat, Giselle, and James H. Stock. 2020. "Quasi-Experimental Estimates of the Transient Climate Response using Observational Data." *Climatic Change* 160:361–71.

Müller, Ulrich K., James H. Stock, and Mark W. Watson. 2019. "An Econometric Model of International Long-Run Growth Dynamics." Working Paper 26593. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26593.

National Academies of Sciences, Engineering, and Medicine (NASEM). 2017. *Valuing Climate Damages: Updating Estimation of the Social Cost of Carbon Dioxide*. Washington: National Academies Press. https://doi.org/10.17226/24651.

Newell, Richard G., William A. Pizer, and Brian C. Prest. 2021. "A Discounting Rule for the Social Cost of Carbon." Working Paper. Washington: Resources for the Future. https://www.rff.org/publications/working-papers/a-discounting-rule-for-the-social-cost-of-carbon/.

Nordhaus, William D. 1992. "The 'DICE' Model: Background and Structure of a Dynamic Integrated Climate-Economy Model of the Economics of Global Warming." Discussion Paper 1009. New Haven, Conn.: Yale University, Cowles Foundation for Research in Economics.

Nordhaus, William D. 2017. "Revisiting the Social Cost of Carbon." *Proceedings of the National Academy of Sciences* 114, no. 7: 1518–23. https://doi.org/10.1073/pnas.1609244114.

Plumer, Brad. 2018. "Trump Put a Low Cost on Carbon Emissions. Here's Why It Matters." *New York Times*, August 23. https://www.nytimes.com/2018/08/23/climate/social-cost-carbon.html.

Ramsey, F. P. 1928. "A Mathematical Theory of Saving." *Economic Journal* 38, no. 152: 543–59. https://doi.org/10.2307/2224098.

Rennert, Kevin, and Cora Kingdon. 2019. *Social Cost of Carbon 101*. Washington: Resources for the Future.

Riahi, Keywan, Detlef P. van Vuuren, Elmar Kriegler, Jae Edmonds, Brian C. O'Neill, Shinichiro Fujimori, and others. 2017. "The Shared Socioeconomic Pathways and Their Energy, Land Use, and Greenhouse Gas Emissions Implications: An Overview." *Global Environmental Change* 42 (January): 153–68.

Rode, Ashwin, Rachel Baker, Tamma A. Carleton, Anthony D'Agostino, Michael T. Delgado, Timothy Foreman, and others. 2021. "Labor Disutility in a Warmer World: The Impact of Climate Change on the Global Workforce." Working Paper. Climate Impact Lab.

Rode, Ashwin, Tamma A. Carleton, Michael Delgado, Michael Greenstone, Trevor Houser, Solomon Hsiang, and others. 2021. "Estimating a Social Cost of Carbon for Global Energy Consumption." *Nature* 598:308–14.

Weitzman, Martin L. 1998. "Why the Far-Distant Future Should Be Discounted at Its Lowest Possible Rate." *Journal of Environmental Economics and Management* 36, no. 3: 201–8. https://doi.org/10.1006/jeem.1998.1052.

COMMENT BY

MAR REGUANT This paper makes significant contributions to updating the social cost of carbon (SCC) calculations, which estimate the marginal damages from climate change and are vital to climate policy (Nordhaus 1982). As explained by the authors, the SCC is often used to inform the design of important legislation, and it has influenced over sixty federal regulatory analyses (Aldy and others 2021).

The authors' goal is to present a framework with several steps to improve current estimations of the SCC. The proposed methodology corrects for the intertemporal changes in the utility of income due to income effects by introducing a stochastic discount factor. The authors also propose eliciting experts' beliefs on uncertain processes, such as emissions, growth, and population trends.

The work by Rennert and colleagues is part of a larger agenda conducted by the Resources for the Future's (RFF) Social Cost of Carbon Initiative and collaborators to improve and expand the tools to inform the metrics surrounding the SCC. This agenda follows the recommendations from a 2017 committee report by the National Academies of Sciences, Engineering, and Medicine (NASEM 2017).[1] This initiative plans to expand further the treatment of uncertainty surrounding economic damages, another recommendation not yet addressed in this contribution which suffers from several limitations.[2]

In this comment, I focus first on the mechanics behind the correction when using a stochastic discount factor. I then discuss the challenges in eliciting beliefs about future outcomes for scenarios that have never happened, a difficulty shared when estimating the SCC with micro data. I also point out the benefits of complementing the SCC calculation with more detailed sectoral analyses. Finally, I conclude by discussing the importance of understanding what the estimates of SCC imply for our abatement strategies, highlighting the necessity of assessing that the recommended SCC is consistent, under reasonable assumptions, with the emissions trajectory that informs it.

1. See, for example, Newell, Pizer, and Prest (2021), which lays out the methodological aspects of the stochastic discount factor in more detail.

2. The treatment of economic damages is at the moment based on Nordhaus (2017), who estimates that by 3 degrees Celsius, the GDP impact will be 2.1 percent. Using the revised DICE model (DICE-2016R), these damages imply an SCC of $31 per ton of CO_2 in 2010 US dollars. These are quite optimistic scenarios, and expanding the set of damages in the presence of uncertainty seems of first order, for example, as in Barnett, Brock, and Hansen (2021).

CORRECTING FOR THE DECLINING MARGINAL UTILITY OF INCOME A key finding of the authors is that considering the declining marginal utility of income, in an intertemporal sense, leads to a lower SCC estimate.[3] Table 1 in the article shows that accounting for stochastic growth discounting (declining marginal utility of income in future events) reduces the SCC by a factor of two to four. The authors find that outliers drive part of the magnitude of the correction. However, muting the role of outliers, the SCC still goes from $96 to $61 when introducing this correction. In sum, considering the declining marginal utility of income leads to less emphasis on climate mitigation efforts.

It is essential to understand the factors driving this finding, which leads to a downward correction of the SCC. This finding clashed with my initial economic intuition, as I had expected the SCC to become more prominent after considering the declining marginal utility of income. My instinct was based on two premises. From a cross-sectional point of view, marginal damages from climate change tend to be largest for poor agents (countries or households), and these agents also have the largest marginal utility of income. From an intertemporal point of view, I expected bad news from climate (severe marginal damages) to correlate with poor growth outcomes, leading to low income and high marginal utility of income as well.

The fact that the SCC is larger for better outcomes is in part driven by the model's assumptions. As explained by the authors, the role of the correction can critically depend on the "climate beta," a parameter summarizing the relationship between damages and aggregate consumption: "The magnitude of this bias [of the SCC that ignores the stochastic discount factor] depends on the climate beta and on the nature of the uncertainty in socioeconomic and emissions trajectories."

The methodology takes a climate beta of essentially one in the short run, partly driven by damages being proportional to GDP (Dietz, Gollier, and Kessler 2018). In good states, marginal damages are *larger* and potentially much larger, given that growth is exponential. However, states with severe damages are also the richest, so they are less important from a relative point of view with Ramsey-like discounting. Therefore, there is a substantial downward correction of the SCC for those states in which the uncorrected SCC is highest.

An additional aspect contributing to a significant downward correction of the SCC is that expert beliefs about growth have a relatively symmetric bell shape centered around a roughly 2 percent growth rate. Because damages

3. Contemporaneous differences in the marginal utility of income are not considered in the paper.

15985-04b_Rennert_Com&GD-6thPgs.indd 296 5/12/22 12:08 PM

are exponential in growth via the previous assumption, marginal damages in high growth states are more heavily corrected downward when compared to low growth states. Still, these are similarly likely, leading to an overall downward correction of the central estimates of the SCC.

It would be helpful to consider departures more explicitly from the baseline assumptions in the model, for example, by assessing the magnitude of the bias correction for a range of climate betas.

ON THE DIFFICULTIES OF FORECASTING Expert beliefs determine trajectories about population, emissions paths, and growth in the proposed calibration of the SCC. However, this is a challenging object to forecast, even by experts, because it is a process that we have never observed.

In truth, forecasting the socioeconomic impacts of climate change is difficult for any methodological approach taken, be it expert elicitation, data estimation, or mathematical modeling.[4] Climate change is an extrapolation of a process that we have never witnessed. The socioeconomic processes behind population growth, emissions, and economic outcomes are extremely difficult to inform in uncharted territory. As soon as we move away from physical processes that follow well-understood laws of physics, nonlinearities in the impacts of climate change are hard to foresee.

The highly uncertain nature of growth, emissions, population, and climate damages affects the precision of the distribution of the SCC. Even without incorporating uncertainty about economic damages, figure 11 shows that there is a wide distribution of implied marginal damages for any given approach or set of assumptions. Therefore, the calculation should be interpreted as an improvement to a "scenario" approach to calibrating the SCC but of a highly uncertain nature.

Whereas this might be the best possible strategy, I believe there is room for improvement in future work. In particular, the modeling of interactions between these three processes is limited, given that the three processes are elicited from different experts. The paper allows for some considerations on the correlations between the distributions, but it is not very clear in the current contribution to what extent they play a significant role. Additionally, it would be interesting to include elicitation of climate betas in future iterations.

The elicitation of independent experts for these categories can lead at times to seemingly inconsistent beliefs. For example, when describing the correlation between growth and emissions, "many [growth] experts expected

4. Pindyck (2013) discusses some of the pitfalls of mathematical assumptions in the calculation of the SCC. Auffhammer (2018) discusses some of the challenges and progress on the empirical data–driven front.

15985-04b_Rennert_Com&GD-6thPgs.indd 297 5/12/22 12:08 PM

that technology breakthroughs in clean energy would dramatically lower global emissions. Implicit in this narrative is a negative correlation between economic growth and carbon dioxide emissions."

However, when looking at emissions elicitation, the beliefs by emissions experts, as implied by figure 7, seem to associate higher emissions to higher growth, at least in the medium run (2050). Some further discussion on how to deal with these relationships could be a potentially fruitful area of improvement.

BRIDGING THE GAP BETWEEN SHADOW VALUES As countries announce pledges to reduce their emissions by 2030 and to achieve net zero by 2050, there has been a focus on how to reduce emissions most effectively for a given goal. This alternative approach calculates the price of carbon that would be consistent with a specific reduction in emissions by solving for the optimal portfolio of climate policies with a given target. The target is taken as given, and therefore this approach is not explicit about what the damages from climate change are. Examples of such an optimization approach are the models solving for how to reach net zero in the United States or the policy framing of the Fit for 55 goals in the European Union.[5]

There has been a recent discussion on the merits and pitfalls of using either approach (Stern and Stiglitz 2021; Aldy and others 2021). While the SCC approach has been preferred in the US context, the target approach is much more common in Europe. This is in part due to the higher agreement on the politically established emissions reduction goals across a broad range of the political spectrum. Europe, indeed, has had a cap-and-trade mechanism with explicit cross-sectoral goals since 2005 and has explicit reduction goals by 2030 and 2050 that are being built into legislation.

While it seems natural to stick to the SCC for US regulatory purposes, I believe there could be some fruitful interactions between these two methods. Like the traditional quantities versus prices trade-off, the two approaches suffer from limitations in the presence of uncertainty about the costs and benefits of fighting climate change (Weitzman 1974). Since the target approach focuses on efficient cost abatement, it tends to have a much more detailed accounting of uncertainties on the technological front, something missing in the highly aggregated integrated assessment models (e.g., Dynamic Integrated Climate Change [DICE], Climate Framework for

5. See, for example, Princeton University Andlinger Center for Energy and the Environment, "The Net-Zero America Project," https://acee.princeton.edu/rapidswitch/projects/net-zero-america-project/; or European Council, "Fit for 55," https://www.consilium.europa.eu/en/policies/green-deal/eu-plan-for-a-green-transition/.

15985-04b_Rennert_Com&GD-6thPgs.indd 298 5/12/22 12:08 PM

Uncertainty, Negotiation and Distribution [FUND], Policy Analysis of the Greenhouse Effect [PAGE]). In the approach proposed by the authors, and given the focus of SCC calculations on damages, such uncertainty is only indirectly present in the trajectories forecasted during expert elicitation.[6]

More accurately describing the expected transformation at a given SCC would add credibility to the forecasted emissions trajectories. Using detailed modeling of technology, even if only for specific sectors in which technology is well understood, such as the electricity sector, could help bound the expected impacts of using the SCC as a tool for cost-benefit analysis. Considering the implications for abatement of a given SCC is particularly important for the proposed methodology, which takes the SCC as an output rather than an equilibrium object, as I discuss next.

A CRITICAL FEEDBACK LOOP In contrast with other attempts at quantifying the SCC, the proposed methodology does not use an integrated assessment model to derive an equilibrium value for the SCC. Instead, the SCC is estimated as the marginal damage of climate change evaluated at the forecasted level of emissions.

This simplicity in the framework can be achieved thanks to directly incorporating expert elicitation for equilibrium trajectories of growth, emissions, and population. However, it comes at the risk of being an out-of-equilibrium object that is potentially inconsistent with the estimated SCC.[7]

The importance of the output from this research, an updated SCC to be used in US policy, makes this concern even more crucial. The SCC is a vital number for US policymaking. As explained in the article, "as political leaders and stakeholders debate both the broad outlines and the fine details of policies to reduce carbon dioxide emissions, the SCC lies in the background as a remarkably important calculation, used by the US federal government for more than a decade for developing vehicle fuel economy standards and power plant emissions rules."

Precisely for this reason, it is important to ensure that the predicted SCC is consistent with the fundamentals that inform it. Estimating damages from climate change cannot be independent from accurately understanding the abatement cost function. Indeed, the equilibrium feedback loop of the calculation can be even more critical if the obtained SCC has a direct impact on climate policy, and thus the emissions trajectory.

6. Experts are considering beliefs on technological progress and breakthroughs when forecasting growth and emissions, but these are not quantitatively modeled.

7. In line with this concern, Diaz and Moore (2017) emphasize the value of using more detailed integrated assessment models that model more closely the functioning of resources and sectors but that ensure that the calculations of the SCC are in equilibrium.

15985-04b_Rennert_Com&GD-6thPgs.indd 299 5/12/22 12:08 PM

To make this point more concrete, it is useful to take the illustrative estimate of $61 SCC in the article. While this is an SCC larger than previously estimated by the US federal government, it is unclear that it can be consistent with the substantial and fast decarbonization efforts that one would need to see to achieve emissions reductions consistent with those expected by the experts. In the electricity sector, one of the most easily decarbonized, it is unclear that such a threshold for cost-benefit analysis would lead to massive transformation and almost net zero generation that many governments have announced.

One response to this criticism is that almost net zero targets in the electricity sector is not a desirable goal, as implied by the estimated SCC. However, because the estimate is not an equilibrium object under the proposed methodology—the heart of this criticism—it is unclear how emissions at this SCC can decrease as projected by the experts. Thus, coupling the estimated SCC with information about expected actions that pass the cost-benefit analysis at such value, even if limited to a subset of policies, would be tremendously useful.

Figure 1 clarifies this point. The increasing cost curve represents the marginal cost curve of abatement. The downward damage curve represents the marginal damages of climate change, which are highest when no abatement occurs and the temperature rises the most.[8] In a simplified environment without uncertainty, the goal should be to equalize the costs of abatement to the avoided damages.

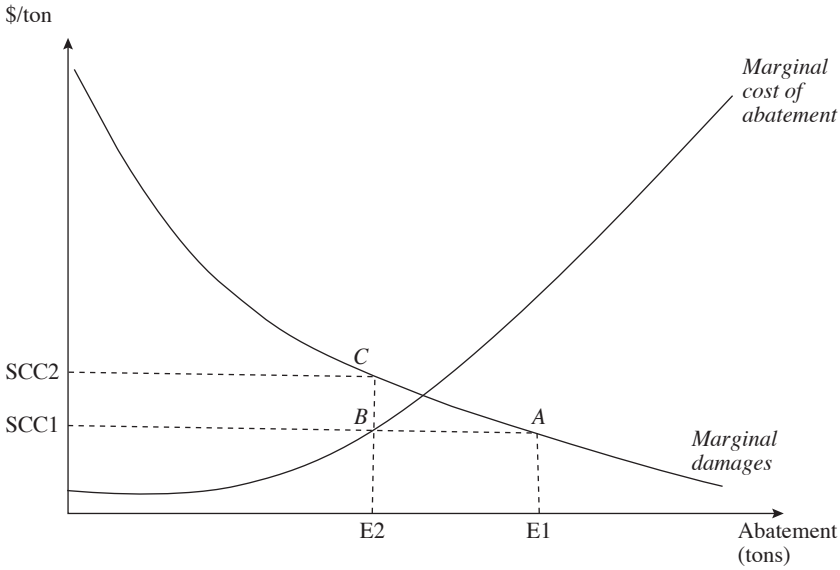
In the proposed methodology, experts are the ones who determine where in the damage function we are crossing. The researcher elicits the emissions reductions from experts and then estimates the SCC at that point. Suppose that experts do not necessarily have equilibrium beliefs about emissions reductions and are optimistic, concluding that abatement efforts will be at point E1. This would imply an SCC of SCC1 (point A). If these beliefs are out of equilibrium, however, abatement efforts might be much lower at a threshold of SCC1, and emissions reductions might only be E2 (point B). However, at E2, marginal damages should have been estimated to be larger and equal to SCC2 (point C).

Any guidance on whether the emissions reductions implied by E1 are plausible under an SCC equal to SCC1 would add validity to the approach. One possibility would be to take a battery of more complex integrated

8. Note that the illustration gives the illusion that this might be a static formulation, but it should be interpreted as a summary of complex relationships with dynamic abatement and damage curves, including stochastic discounting and so forth.

15985-04b_Rennert_Com&GD-6thPgs.indd 300 5/12/22 12:08 PM

Figure 1. Potential Pitfalls with an Out-of-Equilibrium SCC



Source: Author's illustration.

assessment models and run them under the assumption that the benchmark for cost-benefit analysis is SCC1. Because complex integrated assessment models can be sensitive to their technological modeling, the researcher could get a sense of the sectoral and technological assumptions consistent with such an equilibrium outcome.

I believe this is an important missing link. In the example above, I highlight the case where experts might be optimistic about emissions declines because this is the riskiest scenario. Underestimating the need for climate action could lead to substantially more serious marginal damages. Thus, having a more explicit bridge between experts' beliefs and bottom-up approaches that more systematically model expected policies and their impacts on emissions reductions could prove useful to make the recommendations more robust.

CONCLUSIONS The SCC is a critical measure to inform climate policy. The authors present an expanded approach to include stochastic discounting and expert elicitation in the estimation of the SCC. A few limitations are yet present in these calculations. First, it would be important to consider expanded damage functions that are more robust to those based on the DICE

model, something the authors are already exploring. Second, it would be useful to consider complementing the top-down modeling of the SCC with bottom-up approaches of sectoral abatement efforts that give robustness to the estimated SCC.

REFERENCES FOR THE REGUANT COMMENT

Aldy, Joseph E., Matthew J. Kotchen, Robert N. Stavins, and James H. Stock. 2021. "Keep Climate Policy Focused on the Social Cost of Carbon: A Proposed Shift away from the SCC Is Ill Advised." *Science* 373, no. 6557: 850–52. https://www.science.org/doi/10.1126/science.abi7813.

Auffhammer, Maximilian. 2018. "Quantifying Economic Damages from Climate Change." *Journal of Economic Perspectives* 32, no. 4: 33–52. https://doi.org/10.1257/jep.32.4.33.

Barnett, Michael, William Brock, and Lars Peter Hansen. 2021. "Climate Change Uncertainty Spillover in the Macroeconomy." In *NBER Macroeconomics Annual* 36. https://www.nber.org/books-and-chapters/nber-macroeconomics-annual-2021-volume-36/climate-change-uncertainty-spillover-macroeconomy.

Diaz, Delavane B., and Frances Moore. 2017. "Quantifying the Economic Risks of Climate Change." *Nature Climate Change* 7, no. 11: 774–82. https://doi.org/10.1038/nclimate3411.

Dietz, Simon, Christian Gollier, and Louise Kessler. 2018. "The Climate Beta." *Journal of Environmental Economics and Management* 87 (January): 258–74. https://doi.org/10.1016/j.jeem.2017.07.005.

National Academies of Sciences, Engineering, and Medicine (NASEM). 2017. *Valuing Climate Damages: Updating Estimation of the Social Cost of Carbon Dioxide*. Washington: National Academies Press. https://doi.org/10.17226/24651.

Newell, Richard G., William A. Pizer, and Brian C. Prest. 2021. "A Discounting Rule for the Social Cost of Carbon." Working Paper. Washington: Resources for the Future. https://www.rff.org/publications/working-papers/a-discounting-rule-for-the-social-cost-of-carbon/.

Nordhaus, William D. 1982. "How Fast Should We Graze the Global Commons?" *American Economic Review* 72, no. 2: 242–46. https://EconPapers.repec.org/RePEc:aea:aecrev:v:72:y:1982:i:2:p:242-46.

Nordhaus, William D. 2017. "Revisiting the Social Cost of Carbon." *Proceedings of the National Academy of Sciences* 114, no. 7: 1518–23. https://doi.org/10.1073/pnas.1609244114.

Pindyck, Robert S. 2013. "Climate Change Policy: What Do the Models Tell Us?" *Journal of Economic Literature* 51, no. 3: 860–72. https://doi.org/10.1257/jel.51.3.860.

Stern, Nicholas, and Joseph E. Stiglitz. 2021. "The Social Cost of Carbon, Risk, Distribution, Market Failures: An Alternative Approach." Working Paper 28472. Cambridge, Mass.: National Bureau of Economic Research. https://doi.org/10.3386/W28472.

Weitzman, Martin L. 1974. "Prices vs. Quantities." *Review of Economic Studies* 41, no. 4: 477–91. https://doi.org/10.2307/2296698.

15985-04b_Rennert_Com&GD-6thPgs.indd 302 5/12/22 12:08 PM

GENERAL DISCUSSION Michael Kiley offered a comment on emerging literature in the monetary policy space and highlighted the implications of the developing consensus for the paper. Kiley began by noting that while the authors make some efforts to respond to recent data suggesting lower long-run real interest rates, they may have still underreacted to the plausibility of baseline real interest rates below or even well below 2 percent. He argued that estimates of the social cost of carbon (SCC) will be highly sensitive to the prevailing real interest rate, particularly as that rate approaches zero. Given the focus of the paper on the various sources of uncertainty in estimates of the SCC, he suggested that this is yet another source of uncertainty with the possibility of substantially increasing the SCC.

Kiley continued, remarking that the climate beta value is often positive, reflecting assumptions about the damage function and GDP projection. He added that those GDP projection pathways generally reflect fairly stable economic growth over the long run, albeit with considerable uncertainty. Kiley contended that bad climate scenarios are potentially ones in which the macroeconomy is riskier, implying a negative climate beta and a possibly higher SCC. He noted that this is intuitive at the microeconomic level, with increased probability of localized climate disasters resulting in poor economic outcomes, but argued there are possibilities for macroeconomic growth disasters due to warming that increase the welfare impacts of mitigation efforts. Kiley concluded by acknowledging that this would be a new area, but one that is potentially valuable for future iterations of the SCC literature.

David Popp referenced discussant Mar Reguant's concluding point, which identified a feedback loop wherein climate policy itself is an additional source of uncertainty and is impacted directly by calculations of the SCC. Popp wondered whether it was possible for the model to include a projection of the emissions pathway if climate policies followed the proposed SCC and thus could limit the calculation to internally consistent estimates for the SCC (e.g., the SCC is estimated at $56 conditional on some emissions pathway, but that emissions pathway is not itself possible under a $56 cost of carbon). Popp questioned whether there might be some way to link those elements of the calculation to come up with an internally consistent estimate.

Wendy Edelberg pointed out that, due to the wide bands of uncertainty in the paper's SCC estimates, the distributions sometimes suggest a negative value for the SCC. She wondered if that was a natural consequence of simulations that are not bounded at zero or whether that is a legitimate possibility. Edelberg also asked how adaptation plays into the model.

While acknowledging that projecting adaptation into the future is a difficult task, she wondered if that was factored in and whether that may be a contributing factor to the distributions including negative values for the SCC estimates.

Glenn Rudebusch pointed out that there are also cases where the SCC can become infinite, particularly when the real interest rate will be negative for a long time. He noted that this phenomenon has already been evident recently in the data, so this was an additional factor the discounting module would have to grapple with. In particular, Rudebusch questioned how long negative real interest rates could persist, and in turn how far r^* could fall. He referenced his paper with Michael Bauer that attempted to answer a similar question.[1]

Adrian Raftery responded to the questions posed by discussant Reguant and followed up by Popp regarding climate feedback loops within the model. Raftery specified that for the population model, climate feedbacks weren't explicitly included but were implicitly included. He pointed out that it would be difficult to identify exactly what those feedbacks would be, and thus the population model is a statistical model with some adjustments based on expert elicitation. He clarified that the biggest driver of future population is fertility, and there is no evidence to date that climate has a direct impact on fertility (though there is evidence for the inverse). Raftery then mentioned that there are big climate impacts on mortality, but while they result in large damages overall, they are only maybe one-third of 1 percent of total deaths.[2] He argued that small changes in public health would likely have larger overall mortality impacts than the climate. He continued with the point that migration has very little effect on world population overall. One exception is international migration with people from poorer countries moving to relatively richer countries. Raftery posited that climate disasters tend to lead to more internal migration than external migration, with exceptions being small island nations and Bangladesh, meaning that the ultimate effects are hard to nail down.[3] He concluded by arguing that the authors have implicitly included the climate feedbacks

1. Michael D. Bauer and Glenn D. Rudebusch, "The Rising Cost of Climate Change: Evidence from the Bond Market," *Review of Economics and Statistics* (forthcoming), https://doi.org/10.1162/rest_a_01109.

2. World Health Organization, "Climate Change and Health," October 30, 2021, https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health.

3. Shuaizhang Feng, Michael Oppenheimer, and Wolfram Schlenker. 2012. "Climate Change, Crop Yields, and Internal Migration in the United States," working paper 17734 (Cambridge, Mass.: National Bureau of Economic Research, 2012), https://www.nber.org/papers/w17734.

15985-04b_Rennert_Com&GD-6thPgs.indd 304 5/12/22 12:08 PM

toward population because the model is estimated based on seventy years of data from 1950 to 2020, which is a period during which there has been 1 degree Celsius of warming.[4] As such, the model is estimated based on data from a period during which there has been substantial warming and anticipates similar warming rates, meaning the model is implicitly incorporating elements of climate feedbacks.

Kevin Rennert followed, clarifying that the paper does not directly incorporate feedbacks—for example, from climate damages onto economic growth—by design because the intention is for the integrated assessment model to itself calculate the damages and then feed that back onto the pathway for economic growth. This is preferred to experts or the underlying statistical model calculating damages and adding them on top because that would lead to double counting.

Richard Newell spoke to the distinction between the marginal benefits of taking action and the marginal cost of taking action, and the degree to which they should be matched or equated within the model. Newell specified that while integrated assessment models are sometimes used to determine an optimal carbon tax (set at a level in which the emissions level is consistent with marginal damages), the authors intentionally do not do that in this paper. Rather than trying to inform a decision about the global optimal level of emissions, Newell pointed out that their paper intends to aid in the cost-benefit analysis and regulatory analysis for individual regulations or smaller actions in the scheme of global climate action. He contended that it is appropriate to separate and not directly couple emissions within the model to marginal damage estimates.

4. NASA, "World of Change: Global Temperatures," https://earthobservatory.nasa.gov/world-of-change/global-temperatures.

15985-04b_Rennert_Com&GD-6thPgs.indd 306 5/12/22 12:08 PM

RICARDO REIS
London School of Economics

Losing the Inflation Anchor

ABSTRACT Inflation has an anchor in people's expectations of what its long-run value will be. If expectations persistently change, then the anchor is adrift; if they differ from the central bank's target, the anchor is lost. This paper uses data on expectations from market prices, from professional surveys, and from the cross-sectional distribution of household surveys to measure shifts in this anchor. The paper's main application is to the Great Inflation in the United States. The data suggest that the anchor started drifting as early as 1967 and that this could have been spotted well before policymakers noticed it. Other applications using expectations data from Brazil, Turkey, South Africa, the United States in the 1970s, and the United States in 2021 confirm the data's usefulness to measure the inflation anchor in real time.

Where is inflation heading? A natural way to answer this question is to measure where people expect inflation to go. Even if each individual is likely surprised by transitory shocks, there is a general principle that on average, across people and time, they will be collectively right. Therefore, expectations data will reveal persistent changes.

When it comes to inflation, economic theory puts a particularly special role on expectations as its driver. The classical dichotomy dictates that in the long run any inflation rate is consistent with the same real outcomes and welfare. Pinning down inflation (in the literature on determinacy) consists almost entirely of pinning down what people expect inflation will persistently be and how this is influenced by rules, policy regimes, and monetary standards (Castillo-Martinez and Reis 2019). As for the short run, macroeconomic textbooks explain that, if inflation is expected to rise, households will buy more goods today, savers will shift away from nominal assets,

Conflict of Interest Disclosure: Ricardo Reis's research is supported by a five-year grant from the European Research Council. Other than the aforementioned, the author did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this article. He is currently not an officer, director, or board member of any organization with an interest in this article.

15985-05a-Reis-4thPgs.indd 307 5/12/22 12:07 PM

workers will demand higher wages, and firms will post higher prices, all of which lead inflation to rise. The Phillips curve captures the core idea that if expected inflation rises, then only a deep recession can keep inflation down (Mankiw and Reis 2018). It is therefore no surprise that central bankers see it as a key part of their job to keep expectations near the target.

Of course, in equilibrium, both inflation and expected inflation are determined together. Also because of this codetermination, in order to gauge future inflation, beyond data on present inflation, a superior source of information ought to be data on expected inflation. To answer the question of where inflation is heading, we are like a person sitting on the beach, seeing a boat floating with the sea current, trying to figure out if it is drifting away or staying near shore before it is too late. A screen showing the feed from an underwater camera with even a grainy picture of the boat's anchor would be very valuable. This paper tries to build this grainy picture of the inflation anchor.

In particular, I ask whether data on expectations in the past have helped to detect persistent deviations of inflation from target. Looking back at the Great Inflation of the 1970s, could we see signs of what was coming in the expectations data? Do these data help advance the debate on what caused it? Were the same data patterns present in two other experiences of drifting inflation in the past decade, Brazil and Turkey? Or would the data generate false positives, missing the stabilization of US inflation in the 1980s, or the stability of inflation in South Africa in the last decade in spite of successive adverse shocks? Finally, what do the expectations data reveal about inflation in 2021, with all the price volatility brought about by the pandemic?

These are difficult questions to answer empirically. They require having, for the same period in time, both an explosion in inflation and available data on inflation expectations. Yet, while inflation scares were common until the mid-1990s, most reliable measures of inflation expectations start only later. Survey measures in most countries were promoted by central banks under inflation targeting, and this policy regime has so far been successful at keeping inflation stable. As for sovereign bonds whose returns are indexed to inflation, or markets in which one can trade inflation swap contracts, they are also recent and grew in countries where inflation was reasonably under control (Campbell, Shiller, and Viceira 2009). In short, between 1970 and 1995, many anchors were lost, but there are almost no expectations data; between 1995 and 2020, there are data, but no lost anchors.

This paper starts by focusing on the US experience between 1966 and 1973. In the previous seven years, annual inflation had not been above 2 percent for a single month. But, between January 1966 and September 1973,

just before the first oil price shock, inflation averaged 4.3 percent, rising from 1.9 percent to 7.4 percent. It is well appreciated that in the following seven years, from 1974 to 1981, US inflation fluctuated widely and was never below 4.8 percent. But the drift in inflation happened before, gradually and steadily from 1966 onward.

To measure the inflation expectations anchor requires introducing data that have rarely been used. The standard surveys of expectations about Consumer Price Index (CPI) inflation in their current form, the Survey of Professional Forecasters (SPF) and the Michigan survey, start in 1981 and 1978, respectively.[1] Inflation-indexed bonds started only in 1997, and the inflation swap market was only reasonably liquid starting in 2009. Some research has used the Greenbook forecasts, and the mean of earlier vintages of the SPF or the Livingston Survey to assess expectations in the 1960s and 1970s.[2] But their samples are small, making it hard to estimate any cross-sectional moment beyond central tendency.

In this paper I bring in some less-used sources of data on inflation expectations. First, between the second quarter of 1966 and the second quarter of 1977, the Michigan survey included a qualitative question about whether inflation was expected to go up, down, or remain unchanged and supplemented the "up" answers with a follow-up question on by how much, with a choice of five bins. I use those answers to provide measures of cross-sectional mean, variance, and skewness of inflation expectations among households during this key period in the history of US inflation. Second, I use data from the Zurich market on gold forward contracts in US dollars, one of the few international markets where these were freely traded, as I will describe later. Third, I complement these continuous series with more scattered data on expectations by firms and measures of media attention to inflation.

At first sight, it may appear that expectations moved slowly and lagged behind the rise in US inflation. The cross-sectional median (or mean) survey expectation for professionals or households only rose sluggishly and with a lag relative to actual inflation during the seven years. However, between

1. See Federal Reserve Bank of Philadelphia, "Survey of Professional Forecasters," https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/survey-of-professional-forecasters; and University of Michigan, "Surveys of Consumers," https://data.sca.isr.umich.edu/.

2. See Federal Reserve Bank of Philadelphia, "Greenbook Data Sets," https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/greenbook, and "Livingston Survey," https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/livingston-survey.

1968 and 1971, the disagreement captured by the cross-sectional standard deviation and skewness combined with the rising forward prices reveal a clear drift of the inflation anchor. The wage and price controls adopted in 1971 seem to have halted this drift for a little under two years, but after the oil price shock of late 1973, the anchor was clearly lost. Using the approach in Reis (2021) to combine the different expectations data, I estimate a fundamental expected inflation. It shows the anchor drifting by 2.3 percentage points between 1967 and 1971.

In the next section of the paper I look at other episodes, across countries and time, to see whether the expectations data provide a reliable measure of the anchor. First, two cases where inflation drifted away: in Brazil between 2011 and 2016 and in Turkey starting in 2018. In each, looking back today with the benefit of hindsight as to what happened to actual inflation and what were the stances of monetary and fiscal policy, the upward drift of inflation is clear. Looking at expectations data at the time would have signaled that this was the case in real time. Second, I look at a possible false positive, South Africa between 2010 and 2016, which had inflation at the very edge of its target corridor. At the time, it was unclear whether the anchor had been lost or whether an unlucky succession of supply shocks had temporarily raised prices. Today, hindsight points to it being the latter, and again the real-time data on inflation expectations point in the right direction. Third, I look at a case where inflation drifted down, instead of up: the US experience of the early 1980s. The expectations anchor moved accordingly, and again the data in real time reveal it. Finally, I look at the data on expected inflation in 2020 and 2021, where the jury is still out on whether inflation is going to drift away.

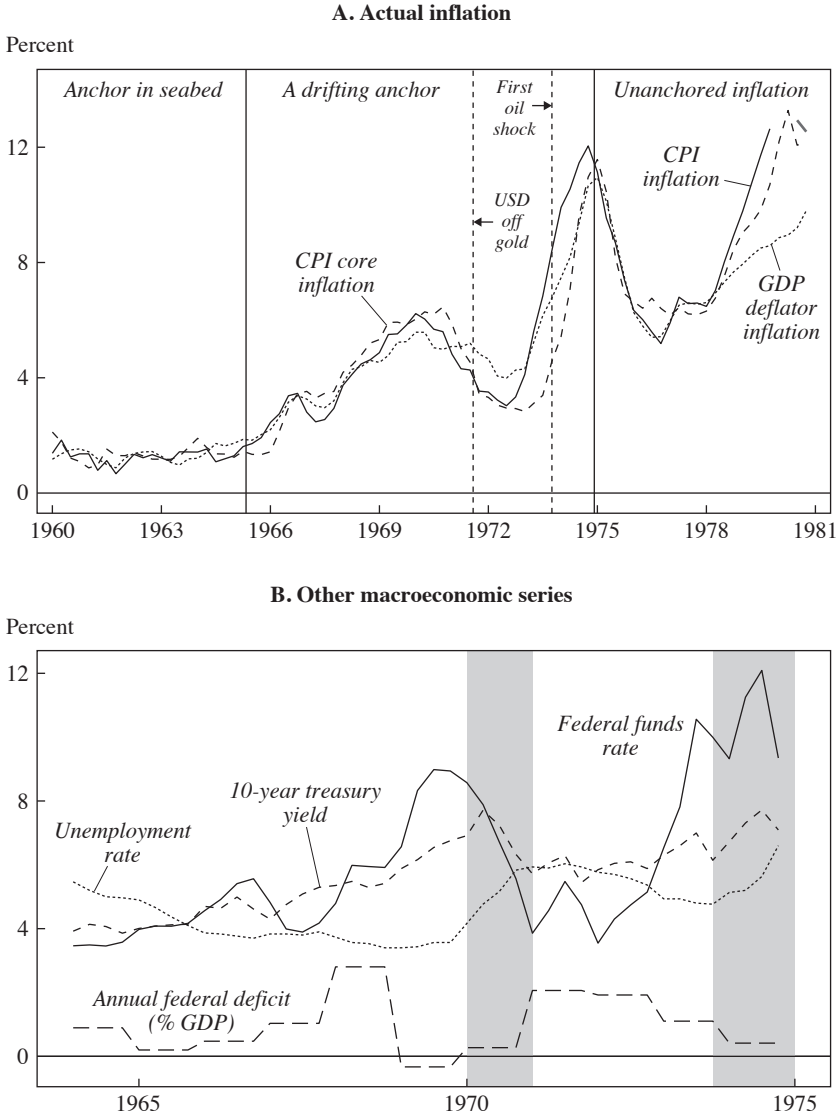
Taken all together, this paper concludes that while market and survey data on expectations are noisy and imperfect, they provide valuable signals on where the expected inflation anchor is.[3]

I. US Inflation 1966–1973: Was the Anchor Adrift?

The top panel of figure 1 plots inflation in the United States during the 1960s and 1970s according to three commonly used price indexes. Before 1965, inflation was remarkably stable, oscillating almost always in a narrow band between 1 percent and 2 percent. The (incomplete and noisy) data on

3. The work closest to this paper has been done by Jalil and Rua (2016), who use news records and scattered business forecasts to infer whether inflation expectations shifted in 1933, triggering the start of the recovery from the Great Depression. See also Binder (2016) for other uses of expectations inflation data in economic history.

Figure 1. The Evolution of the US Economy



Source: FRED.

inflation expectations from this time show an anchor that was firm in the seabed. After the Treasury-Fed Accord of 1951, monetary policy had been focused on inflation control, and an independent central bank was able to deliver it. Fast-forward a decade, to 1975, and there is little doubt that the expectations anchor was lost, with actual inflation being high and volatile. Low and stable inflation returned only after a painful recession in the early 1980s.

The focus of this section is on the decade between the start of 1965 and the end of 1974. Inflation gradually rose between 1965 and 1967, its pace accelerated between 1967 and 1971, and then it slowed down between the middle of 1971 and the end of 1973. At that time, there was a legitimate argument that the anchor was resettling, but as soon as a price shock hit in the form of the rapid rise in oil prices at the end of 1973, it became clear that inflation, whether going up or down, was definitely adrift. There are many discussions of what was behind these movements, which I summarize below, before turning to the focus of this paper: the role that inflation expectations played in policy and academic debates of this time.[4]

I.A. Inflation 1965–1968: All the Signs Were There

INFLATION AND MONETARY POLICY Under the William McChesney Martin Jr. chairmanship of the Board of Governors, the Federal Reserve settled into a policy regime, which he termed "leaning against the wind," that resulted in short-term interest rates moving countercyclically. After a brief recession, the US economy went through an expansion between 1961 and 1965, with the unemployment rate falling by 4 percent but inflation only moderately increasing. Things changed in the second half of 1965. Inflation rose above 2 percent, and the Federal Reserve stepped on the brakes, raising the discount rate in December 1965. It proceeded with other measures to restrict credit in the first half of 1966 that were enhanced in its August 1966 meeting. This policy was a response to both the current state of the economy, as well as to projections of growing military expenditures because of the Vietnam War and growing welfare state spending under President Johnson's Great Society legislation (Meltzer 2005).

However, the monetary restraint was gone by the end of 1966. At the time, the Federal Reserve was accused of causing a credit crunch. The United States was going through turbulent political times, due to social divisions on race and the Vietnam War. Fiscal policy had shifted from being a stabilizer

4. Some excellent discussions are in Blinder (1982), DeLong (1997), Romer and Romer (2002), Meltzer (2010), Nelson (2020), and Hetzel (2020).

of the business cycle in the Keynesian postwar tradition to seeing its role instead as lowering unemployment as much as possible to raise the tide on all groups in the labor market. Martin thought policy coordination with the administration was important, and the Federal Reserve no longer tightened, especially as it subscribed to the view that an unemployment rate above 4 percent justified loose monetary policy.

Not only did the Federal Reserve not tighten further in 1966, in spite of rising inflation (already above 3 percent), but when unemployment slightly rose in the spring of 1967, it reversed course with cuts in interest rates to arrest any chance of a recession. The mild slowdown of 1967 gave way to a strong expansion in 1968, and inflation rose to the 4.5 percent region.

DEBATES IN THE LITERATURE Most accounts of the sources of the Great Inflation find its seeds during this period. A first seed was a view of a static Phillips curve such that any inflation was deemed temporary and acceptable in order to keep unemployment low (Romer and Romer 2002). A second one was the estimates that the economy's acceptable rate of unemployment was 4 percent in spite of a significant deceleration of productivity (Orphanides 2003). Third, there was a misperception of the link between monetary policy and inflation that ignored the role of money supply and did not distinguish between nominal and real interest rates (Nelson 2005). Fourth, despite a permanent and growing external deficit, as well as a large decline in the US stock of gold, the Federal Open Market Committee (FOMC) deferred the defense of the gold peg to the administration and its active use of capital controls, but the Treasury after 1965 subordinated the commitment to the Bretton Woods system to its goals for fiscal spending and low unemployment (Bordo and Eichengreen 2013).

Common to all four factors—the Phillips curve, the natural rate of unemployment, the role of monetary policy, and the defense of the gold peg—was Martin's rejection of explicit economic models and resulting inconsistent framework (Meltzer 2005). Martin gave many speeches opposing inflation and emphasizing its costs, and in his meetings with the Johnson administration he openly disagreed with their sole focus on employment. But, to control inflation, he paid attention to money market data and participants, not economists. Discussions in the FOMC more often focused on current technical market details, especially with regards to the globalization of financial markets and banks' response to capital controls, rather than reflecting on the economic fundamentals of inflation.

THE ROLE OF INFLATION EXPECTATIONS This short-run market-centered view created a special role for inflation expectations. Martin placed a strong weight on investor psychology in financial markets as driving inflation. Early in

his term, testifying to Congress about the slight increase in inflation in the summer of 1959, he stated: "About this time inflationary expectations began to spread. The abrupt upward shift of interest levels in central money markets . . . reflected investor demand for an interest premium to cover the risk of a depreciating purchasing power of invested funds. . . . The experience in the government bond market . . . is a vivid example of the influence of inflationary expectations in financial markets. To the extent that such attitudes come to be reflected in decisions on wages, prices, consumption, and investment, they help to bring about their own realization" (Martin 1959, 113, 119). Later, to explain the clear rise in inflation in 1967–1968, FOMC member speeches frequently referenced "widespread expectations of inflation," and in December of 1968 "the prevailing inflationary psychology" (Federal Reserve Board of Governors 1968, 5, 8). Part of the drift to loose monetary policy during this time can be explained by the subdued values of long-term interest rates. They were used as the measure of both expected inflation and inflation risk premia. Since they were low, Martin inferred that the anchor was firmly in place.

In sum, monetary policy (in relation to inflation expectations) (1) saw the inflation expectations anchor as important to keep inflation stable and (2) inferred expectations from bond rates and market movements; (3) because bond rates remained low, the Federal Reserve continued to be surprised by the drift in inflation in 1965–1968, so (4) it increasingly relied on appeals to unobserved psychology to justify the changes.

I.B. Inflation 1968–1971: The Anchor Comes Loose

INFLATION AND MONETARY POLICY In 1969 and 1970, the economy experienced a mild downturn. GDP fell by less than the fall in federal defense spending. Disposable income actually rose between the peak and the trough of the recession, as the tax surcharge of 1968 expired and the automatic stabilizers in the welfare state kicked in. Monetary aggregates kept rising though, and inflation accelerated to 6 percent. Martin significantly tightened monetary policy; December 1968 is one of the dates Romer and Romer (1989) identified for a contractionary monetary shock. Yet, this arrived late and did not last long. In February 1970, Arthur F. Burns became the new chairman of the Board of Governors. The decision at his first FOMC meeting was to ease monetary policy.

DEBATES IN POLICY AND ACADEMIA At the time, there was an active debate about the Phillips curve, which Burns was keenly aware of. The economists in the recently elected Nixon administration accepted the analysis in Friedman (1968) and thought that any reduction in unemployment achieved

by increasing inflation was temporary (Romer and Romer 2002). But many economists (especially in the previous administration) still thought that an overheated economy would result in higher inflation but not accelerating inflation.[5] On the other side, some academic economists had embraced the view of an expectations-augmented Phillips curve, where the expectations term was proxied by lagged inflation, leading to the so-called accelerationist Phillips curve.[6] By the end of 1971, the latter group had won the debate (in economic circles, even if not in policy circles) in light of the continued rise in inflation in spite of the recession.[7]

Even then, an accelerationist Phillips curve implied that reducing inflation even by a few percentage points back to 2 percent would come with large sacrifices in terms of higher unemployment. Andrew Brimmer, a Federal Reserve Board member between 1966 and 1974, affirmed that there was little doubt within the Federal Reserve: "Fighting inflation, checking inflation was the second priority."[8] Even in April 1969, after all the signs of lower productivity and a rising natural rate of unemployment (Orphanides 2003), Arthur Okun, chair of the Council of Economic Advisers (CEA), wrote that it was compatible to have a long-term "4 percent rate of unemployment and a 2 percent annual rate of price increase" (Okun 1970, 102).

THE ROLE OF INFLATION EXPECTATIONS In the Martin Federal Reserve throughout 1969, there were mentions that "inflationary expectations remained widespread" and of "expectations of continuing inflation" (Romer and Romer 1989, 139). The tightening of monetary policy was justified by appealing directly to expectations: "the Committee agreed that, in light of the persistence of inflationary pressures and expectations, the existing degree of monetary restraint should be continued at present" and that "the Committee took note of the signs of some slowing in the economic expansion and of the indications of stringency in financial markets. In view of the persistence of strong inflationary pressures and expectations, however, the members agreed that a relaxation of the existing degree of monetary restraint would not be appropriate at this time" (Federal Reserve Board of Governors 1969, 121, 145). In July 1969, Martin stated

5. See, for instance, US President and Council of Economic Advisers (1969, 95), written by the departing Johnson administration.

6. However, there was an active debate on whether to interpret the lags on inflation as arising because of adaptive expectations or because of institutional rigidities (see Gordon 1980; Tobin 1980).

7. See Sargent (1999) and Primiceri (2006) for formal models of this change in ideas.

8. Andrew F. Brimmer, interview with Allan H. Meltzer, February 26, 2002, in Meltzer (2005, 169).

that "inflationary psychology remained the main economic problem. . . . It would be a mistake . . . to take any action that might reinforce inflationary expectations just at the time when some weakening in those expectations might be developing" and in August that "it was important for the System not to get into a position of validating the expectations of numerous skeptics who believed the System would ease its policy as soon as it heard the words 'recession' or 'overkill'" (cited in Weise 2012, 37).

With respect to inflation expectations, during the period 1968–1971 (1) they became central in economic discussions as a shifting term in the Phillips curve, but (2) their measurement was still nonexistent beyond using past actual inflation to proxy for it; (3) policymakers kept appealing to shifts in psychology with no hard data to back it.

I.C. From 1971 to 1974: The Anchor Adrift

INFLATION AND MONETARY POLICY In 1972, the economy turned sharply to a strong expansion. Real GDP grew by an astounding 6.9 percent by the end of the year (but the unemployment rate fell by only 0.8 percent). Policy was strongly procyclical, as the decline in government purchases of the previous years was halted, and the Federal Reserve kept policy loose. From the perspective of inflation though, two milestone events in 1971 drove a fall in CPI inflation during that year, and an acceleration in the next.

The first of these was the end of Bretton Woods, with the suspension of dollar gold convertibility on August 15, 1971. The steady increase in inflation in the previous six years had caused a large outflow of gold from the United States, and such a large imbalance in the balance of payments was no longer sustainable. Without this constraint, inflation was free to rise much more, and the dollar depreciated continuously between 1971 and 1973. Expectations of this depreciation could now drive inflation.[9]

The second key event was the imposition of a wage and price freeze by the Nixon administration on August 15, 1971, as part of its New Economic Policy (NEP). This brute-force approach to keep CPI inflation down was fully supported by Burns. Nominal wage growth had risen by more than 8 percent per year leading up to 1970, in spite of the recession and stagnant productivity, and he blamed this on the monopoly power of labor unions and the generosity of welfare programs. An automobile labor strike in 1971

9. By itself, the move from the (adjustable) pegs of Bretton Woods to flexible exchange rates did not have to cause inflation. Free from the influence of US inflation, West Germany went through a disinflation in 1971, almost a decade before the United States had its own.

cemented this view, and the annual report of the Federal Reserve was filled with analyses of unit labor costs. Burns dismissed the fast growth in monetary aggregates and thought that the controls and other guideposts would break the expectations of workers. Hetzel (1998) argues that, even well before this episode, Burns had long regarded wage-price spirals driven by the expectation of inflation as a key driver of high inflation: "One of the main factors in the inflation that we have had since the end of World War II is that many consumers, businessmen, and trade union leaders expected prices to rise and therefore acted in ways that helped to bring about this result" (Burns 1957, 71). Government interference in private price and wage setting was to be the way of breaking the spiral.

From then on, between 1972 and 1974, inflation accelerated. Besides the bounce-back from the price controls and the continued effect of the depreciation of the dollar, the most well-known direct cause was the rapid increase in oil and food price shocks. Crop failures in 1972, an El Niño event that made Peruvian anchovies disappear, and large sales of US wheat to the Soviet Union led to a surge in world food prices in 1973–1974. The OPEC oil embargo after the 1973 Yom Kippur War led to oil prices increasing by a factor of three by 1974. By themselves, they can potentially explain more than half of the increase in inflation and its fall in the following years once oil prices stayed high but stopped rising (Blinder and Newton 1981; Blinder and Rudd 2013).

Yet, the reversal down of inflation after the oil shocks was far from complete. A possible explanation is that the temporary inflation shock raised inflation expectations becoming inertial. Policymakers at the time thought so, leading to the famous Whip Inflation Now (WIN) public relations campaign in 1974 to try to lower the psychology of inflation expectations, a famous failure of policy communication that is not backed by actual policy actions.

THEORIES AND EXPECTATIONS Beyond supply shocks, the literature has offered two other factors that contributed to the acceleration of inflation in these years. Each involves expectations at their center.

The first of these was a confusion between transitory and permanent shocks at the Federal Reserve. At the helm, Burns was confident of his forecasting ability guided by reduced-form empirical analysis of data, shaped by his experience as president of the National Bureau of Economic Research (NBER) and its measurement without theory tradition. In his words, testifying to Congress in 1974, "the current inflation began in the middle 1960s when our government embarked on a highly expansive fiscal policy" (Hetzel 1998, 37). In turn, he thought the inflation of 1973 was due to food

and oil prices, and the further increase in 1974 was due again to budget deficits (even though those had been small). There was always a temporary shock to explain the persistent drift. Meltzer (2005, 160) describes this period as one when, among the Federal Reserve staff, "they gave special explanations—a relative price theory of the general price level—in effect claiming that the rise in the price level resulted from one-time, transitory changes that they did not expect to repeat." Likewise, by the end of 1972, an unemployment rate of 5 percent was seen as close to full employment, higher than the previous 4 percent but still well below what our estimates today suggest (Reis 2003). This sluggish adjustment upward of the natural rate in spite of the higher unemployment and acceleration of inflation of previous years implicitly reveals a view that this was due to temporary shocks to the natural rate of unemployment, coupled with a belief that expected inflation had not changed.

Burns viewed monetary policy as affecting credit conditions and through it the confidence and investment proclivity of businessmen, with money at best reflecting these rather than causing them (Hetzel 1998). Shifts in expectations were then always tendentially transitory, with a multitude of policies being able to bring them back in line. For instance, testifying to Congress in 1964 about deficit reduction, he stated: "In this new psychological environment, our trade unions may not push quite so hard for a large increase in wage rates, since they would no longer be anticipating a higher inflation rate. And in this new psychological environment, our business people would not agree to large wage increases quite so quickly" (Hetzel 1998, 38).

This last quote points to the second factor: the emergence of an inflation bias through expectations. As articulated by Tobin (1980), this was a break with the Keynesian thought of the previous two decades, since it implied that monetary and fiscal policy management would be unable to cure stagflation. Arthur Burns (1987, 695) described this in his 1979 lecture that reflected on his experience: "In earlier times, when a central bank permitted excessive creation of money and credit in times of prosperity, the price level would indeed tend to rise. But the resulting inflation was confined to the expansion phase of the business cycle; it did not persist or gather force beyond that phase. Therefore, people generally took it for granted that the advance of prices would be followed by a decline once a business recession got under way. That is no longer the case. Nowadays, businessmen, farmers, bankers, trade union leaders, factory workers, and housewives generally proceed on the expectation that inflation will continue in the future, whether economic activity is booming or receding."

15985-05a-Reis-4thPgs.indd 318 5/12/22 12:07 PM

One of the reasons for this change was the political influence over monetary policy. Especially through 1972, the Nixon administration, and the president personally, strongly pressured the Federal Reserve not to lean against the economic expansion and jeopardize the fall reelection, and the minutes and transcripts of the Federal Reserve show a concern about the political repercussions of its measures (Weise 2012). However, they also reflect a view that this would not compromise long-run price stability because the election would be over soon. The inflation bias that yielding to political pressure would have on expectations given recurrent elections was dismissed, and yet it was visible in the data for the two decades that followed (Alesina 1988). As Gordon (1980, 145) writes: "The overheated expansion of 1972–73 is perhaps the leading postwar example of Nordhaus's 'political business cycle' in action."

In short, while the oil price shocks were significant and unforeseen, the shoot up of inflation in 1971–1974 and its lasting effects through the rest of the decade worked in part through their effect on expectations since (1) the gold peg was not replaced by another regime that would keep the expectations anchor steady, (2) hope was placed on price and wage controls affecting inflation expectations, (3) communication—not backed by actual policies—failed to bring down inflation expectations, and (4) internal forecasts that all shocks had temporary effects led to a belief that inflation expectations remained anchored, even as there were growing suspicions of an inflation bias in expectations emerging, but no data confirmed either.

I.D. Conclusion: Expectations as Add-Factors

The decade between 1965 and 1974 was filled with shocks, changes in regime, and contributing causes to the rise in inflation. The literature that has inspected and debated them has fallen into three competing hypotheses (Romer 2005): a politics view that focuses on the pressure from expansionary fiscal policy and the lack of independence by Martin and Burns relative to the presidential administrations; the ideas hypothesis which focuses on incorrect views of the Phillips curve, the value of the natural rate of unemployment, the power of monetary policy, and the role of Bretton Woods; and the supply shocks hypothesis which focuses on declining productivity, the Nixon price controls, and the OPEC oil shock. Across all of them, the political context of the time played an important role, and debates on the relative weight of each view will continue.

This section highlights that expectations play an important role in these discussions. Policymakers used changes in expectations to justify why policies had failed in the past but would not in the future. Academics appealed

to shifts in expectations to explain why economic relationships changed, why the monetary policy regime led to high inflation, or why inflation stayed elevated. Throughout, however, data on expectations were conspicuously missing. Like add-factors in forecasting models, accounts of the Great Inflation use private inflation expectations as unobservables that help explain some of the data.

At times, data on expectations were replaced by appeals to the psychology of markets or businessmen. Other times, they were only indirectly inferred from changes in macroeconomic variables using particular economic theories. With the widespread use of rational expectations from the 1980s onward, expectations became tightly tied to observed fundamentals, so the obstacles in measuring and observing them were circumvented altogether. None of these options are very useful when trying to assess whether inflation expectations are anchored in the present. Appeals to how people may be feeling are too subjective. Macroeconomic variables can only be used with a long delay to infer what expectations were in the past. And, while expectations are undeniably forward-looking, they are also not strictly rational in the rational expectations sense, leaving us with a wide set of models of where they may be drifting. The next section turns to real-time data from surveys and market prices to see whether they can provide timely signals of how solid the anchor is.

II. Data on Inflation Expectations

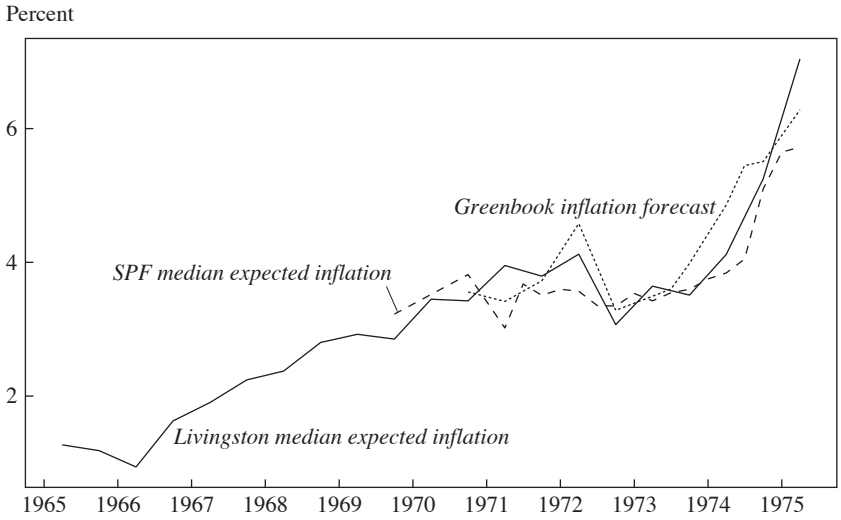
This section canvasses sources of data on the inflation expectations anchor during the period when inflation went adrift.

II.A. Inflation Expectations: Professionals

The first source of data was surely looked at by policymakers: the forecasts made by the professionals who work at the Federal Reserve System. These are published in the Greenbook and are nowadays made public by the Federal Reserve Bank (FRB) of Philadelphia. Starting in 1967:Q1, the Federal Reserve staff produced forecasts for the GDP deflator over the next three months. Starting in 1969:Q4, there are also forecasts over the next twelve months, and so for annual inflation one year ahead.[10]

10. Federal Reserve Bank of Philadelphia, "Greenbook Data Sets," https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/greenbook.

Figure 2. Inflation Expected by Professionals

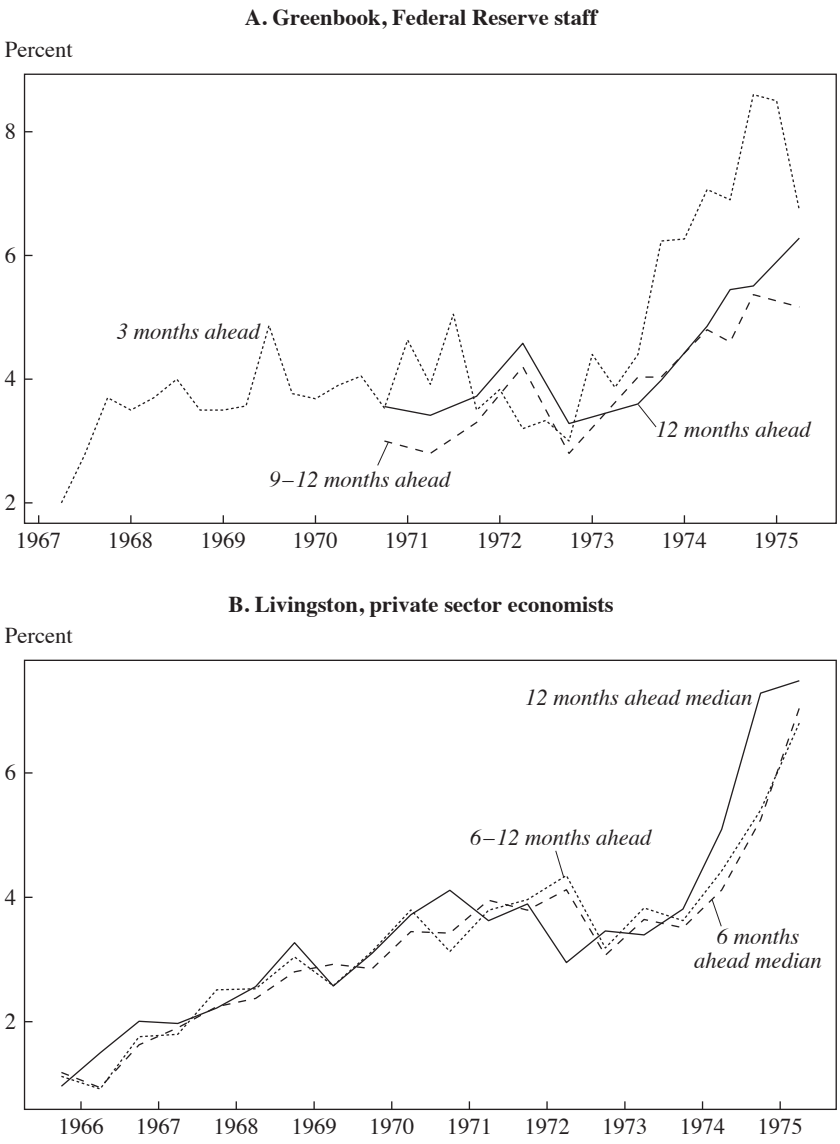


Source: Mankiw, Reis, and Wolfers (2004), from Federal Reserve Bank of Philadelphia (https://www.philadelphiafed.org/surveys-and-data/real-time-data-research).

The Greenbook inflation forecasts are plotted in figure 2 where the date at which the forecast was made is on the horizontal axis (so in quarter t it shows the report made at t for inflation between t and $t + 4$). The data are consistent with the narrative analysis of Federal Reserve statements in section I.C that the Federal Reserve thought the successive rises in inflation were temporary one-off deviations. The forecasts were significantly below the realizations of inflation and only adjusted upward sluggishly. Panel A of figure 3 confirms this view by plotting the expectation three months ahead, twelve months ahead, and nine to twelve months ahead. Inflation was almost always expected to be higher in the nearer term than in the term farther away.

At the same time as the Greenbook forecasts became available, the American Statistical Association and the NBER started surveying a group of professional forecasters working for major banks and large corporations. The typical survey at the time had ten to thirty respondents and took place every quarter. The series in these data for twelve-month ahead CPI inflation forecasts is commonly used in the literature on inflation expectations, but it only starts in 1981:Q3. However, the survey started earlier, and from 1968:Q4 on, it included questions on the expected growth of nominal and real GDP. Combining the two yields the professionals' forecasts for the

Figure 3. Expectations of Inflation at Different Horizons



Source: Mankiw, Reis, and Wolfers (2004), from Federal Reserve Bank of Philadelphia (https://www.philadelphiafed.org/surveys-and-data/real-time-data-research).

GDP deflator.[11] A less used but similar survey of professionals was undertaken by journalist Joseph Livingston who asked between twelve and forty-eight economists in academia, business, and markets, twice per year, starting in 1946, about CPI inflation over the next six months and over the next twelve months.[12]

The cross-sectional median in the two surveys of expected inflation over the next twelve months is plotted in figure 2. Because the size of the samples is small, one cannot reliably estimate the cross-sectional distribution beyond a first moment. The median forecast, similar to the Federal Reserve's forecasts, was sluggish and slow to keep up with the rapid increase in inflation at the time. There is a telling difference between the two revealed by comparing the two panels of figure 3. The private professionals seemed less convinced that the inflation rise was temporary, as they would raise their expectations at the different horizons in lockstep.

II.B. Inflation Expectations: Households

The main survey of household expectations of inflation is the Survey of Consumer Attitudes and Behavior administered by the University of Michigan Survey Research Center (SRC).[13] Since January 1978, it has surveyed around 500 to 700 households every month, asking them: "By about what percent do you expect prices to go up, on the average, during the next twelve months?" An early version of the quantitative inflation question was tried in 1976 and 1977, only every six months, still outside our period of interest.

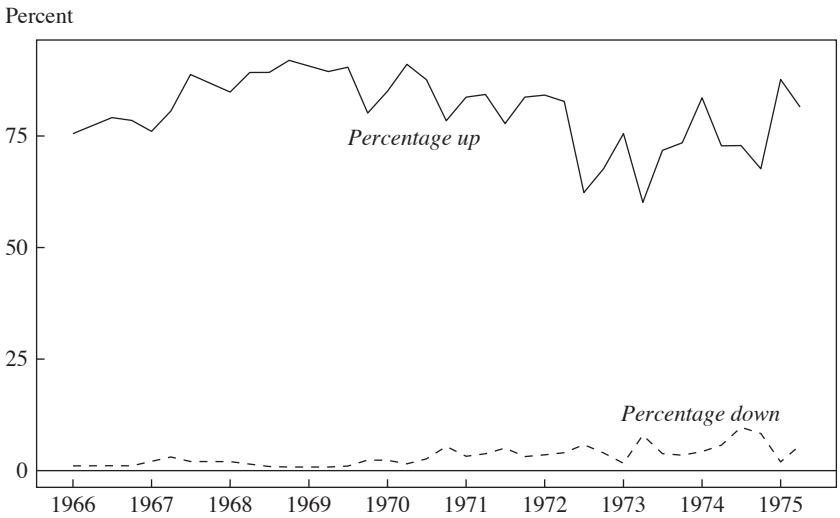
Before 1976, and starting in 1946, the survey existed but it was quarterly and it only asked a qualitative question about inflation. The precise question changed over time, but between 1961:Q2 and 1977:Q2 it was: "Speaking of prices in general, I mean the prices of the things you buy—do you think they will go up in the next year or go down?" Figure 4 shows the time series of the answers to that question between 1965 and 1974. It is hard to see much of a change. At the same time, with the majority of respondents already anticipating prices to rise from the start, a shift up in how much they think so would not be detected.

11. Federal Reserve Bank of Philadelphia, "Survey of Professional Forecasters," https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/survey-of-professional-forecasters.

12. Federal Reserve Bank of Philadelphia, "Livingston Survey," https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/livingston-survey.

13. University of Michigan, "Surveys of Consumers," https://www.src.isr.umich.edu/projects/surveys-of-consumers-attitudes-sca/.

Figure 4. Household Qualitative Expectations of Inflation Going Up or Down



Source: Mankiw, Reis, and Wolfers (2004), from Survey of Consumer Attitudes and Behavior, University of Michigan Survey Research Center (https://www.src.isr.umich.edu/projects/surveys-of-consumers-attitudes-sca/).

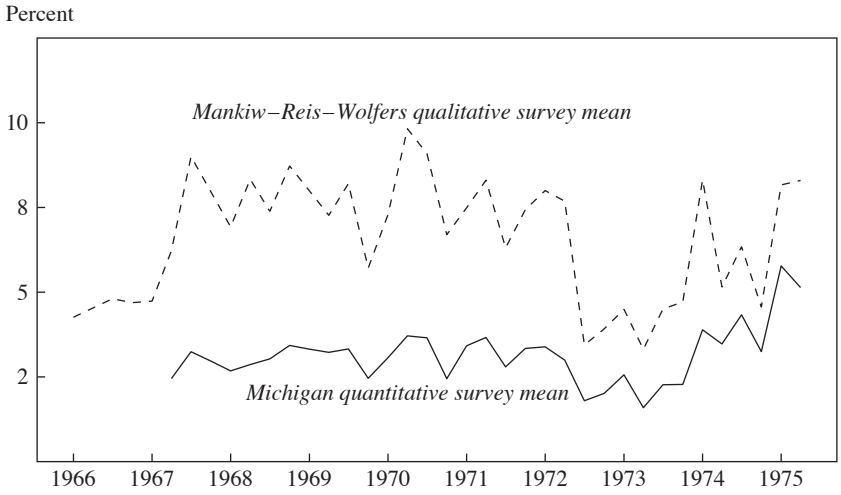
Luckily, the SRC asked a supplementary quantitative question after this one between 1966:Q2 and 1976:Q4. Respondents who expected prices to rise were asked to state in which of a few bins their mean expectation would be. The precise question was: "How large a price increase do you expect? Of course, nobody can know for sure, but would you say that a year from now prices will be about 1 percent or 2 percent higher, or 5 percent higher, or closer to 10 percent higher than now or what?" While the answers to this question were used in academic articles published in the early 1980s, the data seem to have been forgotten (and they are not easy to access); I found no paper using them that was published in the last twenty years. To convert the bin answers into numerical answers, I use the standard (but imperfect) procedures in the literature (Pesaran and Weale 2006).[14]

Figure 5 shows the resulting cross-sectional mean of expected inflation. Also plotted in the figure is a series built by Mankiw, Reis, and Wolfers (2004) that used only the answers to the up or down question and assumed

14. My estimates are not too different from those in Juster and others (1972) and Juster and Comment (1980).

15985-05a-Reis-4thPgs.indd 324 5/12/22 12:07 PM

Figure 5. Cross-Sectional Mean of Expected Inflation by Households



Source: Mankiw, Reis, and Wolfers (2004), from Survey of Consumer Attitudes and Behavior, University of Michigan Survey Research Center (https://www.src.isr.umich.edu/projects/surveys-of-consumers-attitudes-sca/).

a normal distribution to back out a mean and a standard deviation at every date. As with the first moment from the professional surveys, there is no discernible trend in the central tendency of the household survey throughout the decade. As the literature has found for other periods, also during this time, average household expectations appear very sticky (Coibion and Gorodnichenko 2012).

II.C. Disagreement and the Cross-Sectional Distribution of Expectations

In light of what we know today from the literature on expectations, it is not so surprising that measures of the first cross-sectional moment of household surveys are seemingly unresponsive to shocks and only very sluggishly adapt to new regimes (Coibion, Gorodnichenko, and Kamdar 2018). Most people are uninformed, inattentive, or just ignorant about inflation. However, the literature that has focused on the data since the 1980s has also found that the cross-sectional distribution of expectations contains plenty more information (Mankiw, Reis, and Wolfers 2004; Coibion and Gorodnichenko 2012). Modern theories of inattention, learning, or mental biases highlight that people differ in how fast they adjust to news in systematic ways (Mankiw and Reis 2010; Angeletos, Huo, and Sastry 2020).

By comparing their expectations, we can infer where they are eventually heading, and thus where the anchor is. The distribution captures disagreement between people, some better informed than others. It can be quite informative of what the anchor is.

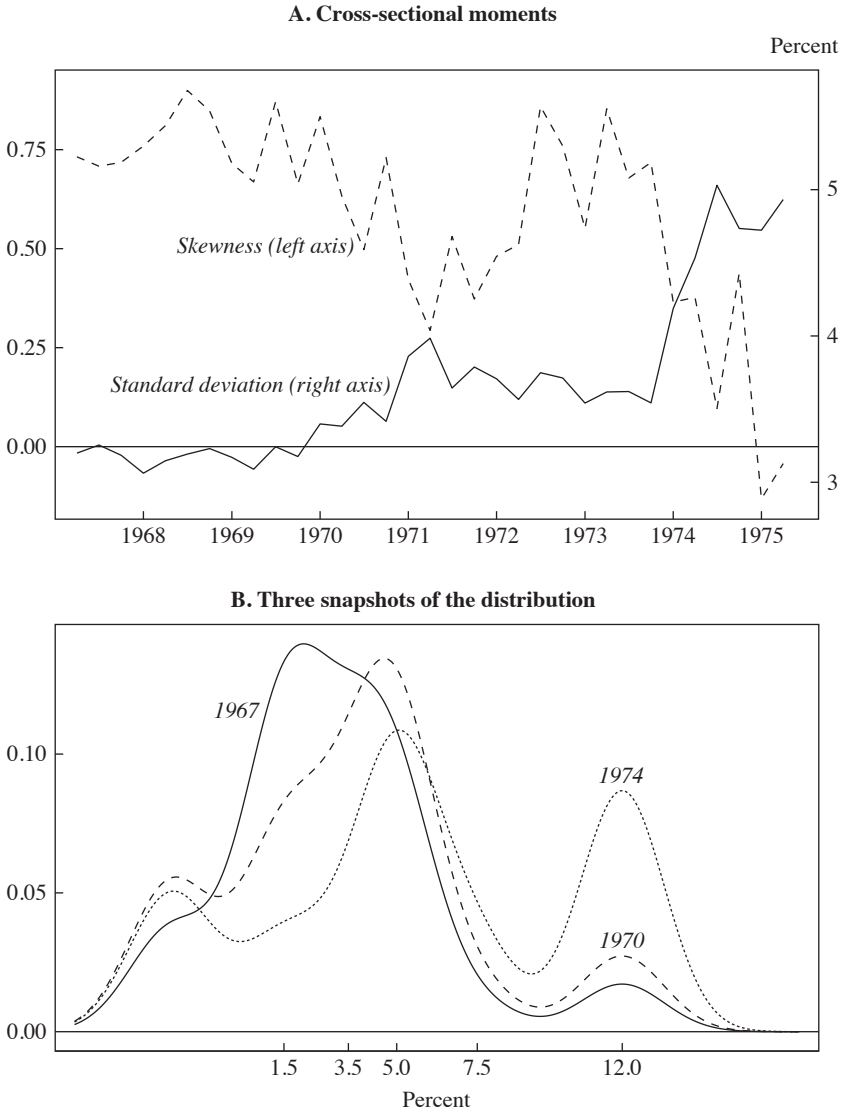
Panel A of figure 6 plots the cross-sectional standard deviation and skewness of household expectations using the Michigan survey. Starting with the forecasts from 1968 about inflation in 1969, disagreement in the series clearly starts rising for about two years. For the following three years, disagreement is approximately unchanged. But after the oil shock at the end of 1973, disagreement about expected 1974 inflation jumps up and stays elevated after that. Skewness, which throughout is positive, also changes around 1968, clearly falling for the first two years. When disagreement settles, skewness starts rising, and when at the end of 1973 disagreement jumps up, skewness jumps down.[15]

Combined, these two series tell the following story, which panel B with the cross-sectional distribution at three points in time confirms. While inflation may have started drifting up in 1965, it is only in the forecasts made around 1967–1968 that expectations seem to show some drift in the anchor. Section I.B discussed how the 1969–1970 downturn marked a turning point in academic thought, with its clear rejection of a static Phillips curve as inflation accelerated. In household expectations, there is also a difference. The mass of agents expecting low inflation thinned out as more households moved to the long right tail of the positively skewed distribution. This raised dispersion while lowering skewness. Then, between 1971 and the middle of 1973, the incomes policy seems to have worked in the way Burns hypothesized, by reversing the increase in expected inflation. This is consistent with the robust empirical finding that people form their expectations using a subset of goods and that large changes in public policy trigger attention to inflation (Candia, Coibion, and Gorodnichenko 2020).

Yet, the oil shock, which had only a small and delayed effect on the first cross-sectional moment, had a large impact on the second and third moments in 1974. There was an immediate large movement of mass from the left and middle of the distribution to its right tail. If the wage and price controls had an effect on expected inflation, it was only a temporary one.

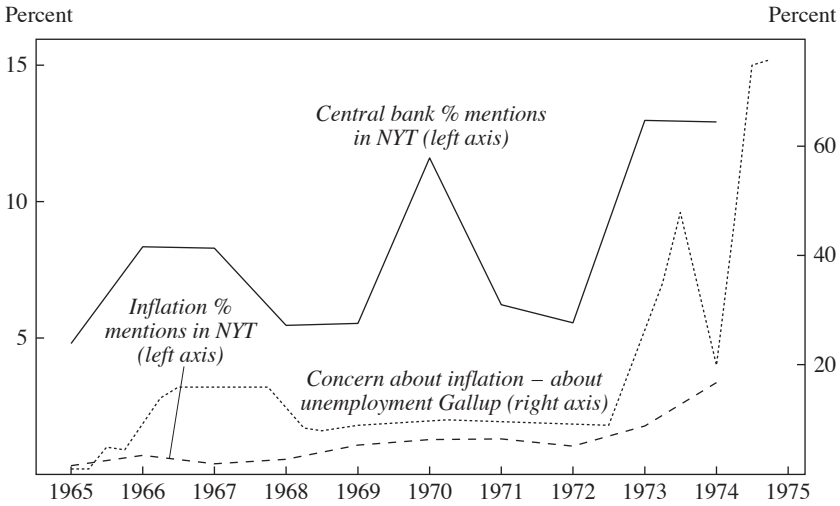
15. Estimates of kurtosis or higher-order moments show no appreciable trend. This is also the case in the quantitative Michigan survey answers after 1978: three cross-sectional moments seem to suffice to capture all the relevant time-series variation in the distribution.

Figure 6. Disagreement about Expected Inflation among Households



Sources: Mankiw, Reis, and Wolfers (2004); Survey of Consumer Attitudes and Behavior, University of Michigan Survey Research Center (https://www.src.isr.umich.edu/projects/surveys-of-consumers-attitudes-sca/).

Figure 7. Media Mentions and Public Concerns Related to Inflation



Source: Author's calculations, using ProQuest database and Fischer and Huizinga (1982).

II.D. Other, Imperfect, Proxies for People's Expectations

Figure 7 plots a proxy for attention to inflation news by households: the number of news references to inflation or to the Federal Reserve, in articles published in the *New York Times*. Some research suggests that the news can proxy for people updating their expectations more often (Carroll 2003). The uptick in news stories mentioning inflation starts in 1968. A relative decline in mentions in 1972, and the surge upward after the oil price shock, tracks the patterns of disagreement in the household survey.

The figure also shows the difference between the share of households that report that inflation is the nation's biggest problem minus those that mention unemployment as a concern instead in the Gallup poll. The uptick here is higher, and it happens already at the end of 1966, with no decline in 1970–1972. The rise after the oil shock is again considerable.

One concern with the survey data used so far is that their horizon was for the most part one year. Yet economic theory would suggest that it is longer-horizon inflation expectations that provide the anchor for the persistent component of inflation. Moreover, insofar as inflation may sometimes be affected by shocks with a half-life of two or three years, then one-year-ahead expectations may be more volatile than their five- or ten-year-ahead counterparts, giving a misleading impression of an anchor that

15985-05a-Reis-4thPgs.indd 328 5/12/22 12:07 PM

is not firm. The Michigan survey started asking regularly, every month, about long-term inflation expectations after 1990. Focusing on the decades of the 2010s, the correlation between the interquartile range of the one-year-ahead and the long-term inflation answers is 0.48, and the correlation between the skewness measures is 0.39. There is considerable month-to-month noise in estimating these correlations, very visible in figure 6, but the low-frequency movements that were emphasized are quite similar.

Finally, because firms choose prices, the expectations of their managers may be particularly important as an anchor. Most of them will not employ a professional economist, but at the same time they may have a greater financial incentive than households to get their forecasts right. However, research has found that firms do not seem to make more accurate forecasts than households, and too often they confuse industry-specific relative price changes for overall inflation (Candia, Coibion, and Gorodnichenko 2021; Andrade and others 2020).

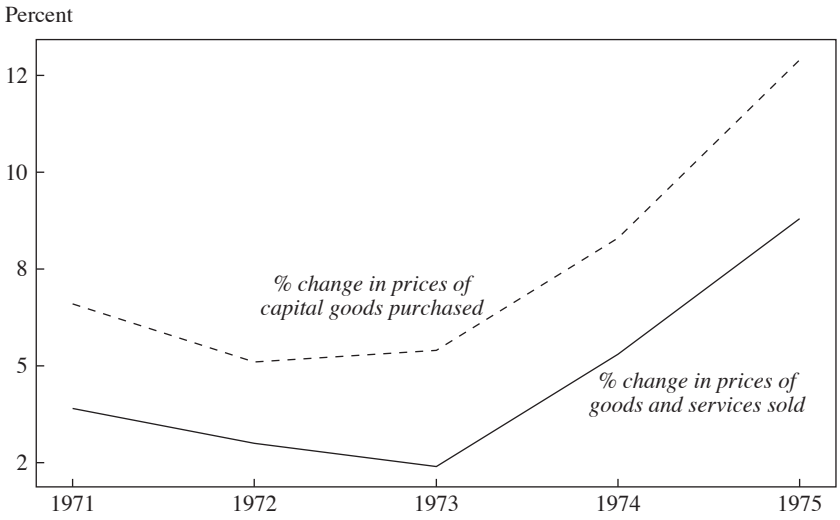
The only series that I was able to find on firm expectations was by the Bureau of Economic Analysis as part of the annual survey of business expenditures on plant and equipment published in the Survey of Current Business from 1970 onward. Approximately 3,200 firms answered the question in year t: "What are your best estimates of average price changes from $t-1$ to t and from t to $t+1$? (Approximations are acceptable.)" The question was asked separately for capital goods purchases and for the goods and services sold; figure 8 reports both, as tabulated by de Leeuw and McKelvey (1981). The pattern in the figure roughly confirms the fact that by 1971 expectations were elevated, but over the next two years they seemed to fall before shooting up in 1974 following the oil price shock.

II.E. Inflation Expectations: Markets

Most central banks today look at market prices to infer inflation expectations at high frequency. These can be useful because of the strong financial incentive that traders have to get this right. Moreover, the market price reflects the expectation of the marginal trader. Comparing it with that of the average trader, or person, provides a signal of the distribution of inflation expectations. At the same time, market prices are often polluted by noise arising from financial and liquidity frictions.

Market expected inflation is usually measured using either the price of inflation swap contracts or the difference between yields on inflation-indexed and nominal government bonds. Neither existed in 1965–1975; McCulloch (1980) argued that these contracts could not exist because they were illegal. In 1933, Congress had approved a "gold clause" prohibiting

Figure 8. Firm Expectations of Inflation



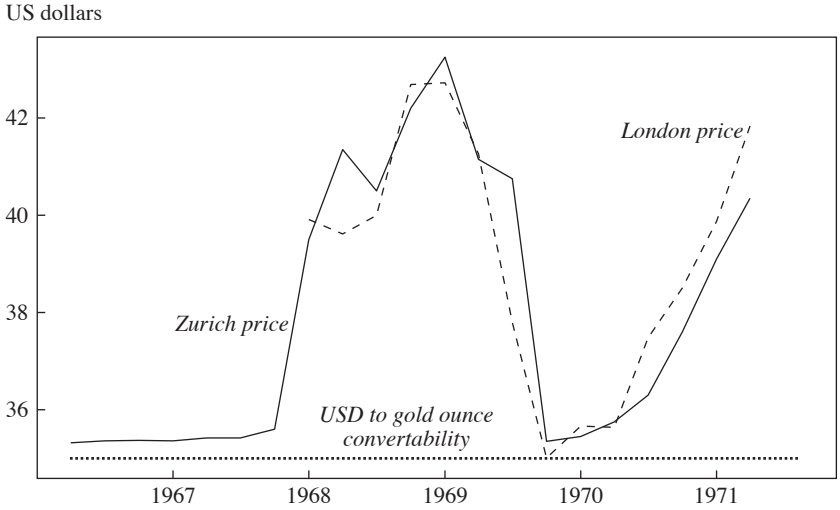
Source: De Leeuw and McKelvey (1981).

bonds indexed to the price of gold as being against public policy. Under Bretton Woods, court rulings extended this prohibition to any domestic contract indexed to the price level. The clause wasn't explicitly repealed until 1977, at which point futures contracts for gold started being traded in the United States.

The US Treasury issued foreign currency denominated bonds in West German marks, Swiss francs, Italian lira, Belgian francs, and Austrian schillings. These "Roosa bonds" were medium-term notes issued between 1962 and 1971 to help relieve the pressure on US gold holdings during the Treasury's attempt to defend the Bretton Woods system in spite of rising US inflation. But they were non-marketable, held by foreign governments. Their interest rate does not reflect market expectations.

Chairman Martin often referred to long-term bond yields as a measure of market expected inflation. If real interest rates and long-term interest rate risk were constant, then these would be market forecasts of the persistent component of inflation. However, during this decade, the argument that the natural rate of unemployment and productivity growth changed significantly would imply with equal force that the real rate would have changed as well. Moreover, the literature on the term structure of interest rates almost always finds that risk premia move around considerably

Figure 9. Gold Prices, Spot, per Ounce



Sources: Pick (1971); FRED.

(Ang, Bekaert, and Wei 2007). An alternative used by Friedman and Schwartz (1963) is the equity premium, but again today it is understood that it is more likely to reflect changes in risk premia than in expected inflation (Campbell, Sunderam, and Viceira 2017).

A final dead end is the US dollar price of gold that was traded freely on the London Mercantile Exchange. The Federal Reserve and other foreign central banks formed a gold pool, combining their stocks of gold to intervene in that market to keep the price of an ounce of gold close to its peg of $35. The pool failed in March 1968, the London market temporarily closed, and even after it reopened, there were US capital controls and regulations of the London market. Perhaps as a result, there was no trading of organized gold futures in the London market until 1975.

Across the channel, in Zurich, there had been a spot gold market that was just as active since the end of World War II. There were no capital controls in Switzerland, no legal restrictions on writing gold contracts, and no limits to holding, exporting, or importing gold. In the immediate postwar, Zurich was the major gold trading center, only losing to London later when the market in London opened in 1954. The market boomed with activity during the gold pool in the London market, and there was trading in multiple currencies. American residents and corporations became large investors in the late 1960s. Figure 9 plots the spot US dollar price of a gold ounce in

the London and Zurich markets during these times. They are very close to confirming that the Zurich market price signals were arguably as informative as those from the large London market that are commonly used by researchers.

These markets arose following the London gold bubble of 1960. During the Bretton Woods system, the loss of an inflation anchor had to be tied to leaving convertibility with gold. Therefore, the US dollar prices of future contracts of gold delivery would provide an indicator of expected inflation. *Pick's Currency Yearbook* for 1971 reports representative annual premia in these contracts throughout the 1960s. The implicit expected devaluation of the US dollar relative to gold can be taken as a proxy for market expected inflation. Between 1966 and 1971, this implicit market expected inflation was 1.4 percent, 1.8 percent, 3 percent, 4.4 percent, 3.1 percent, and 4.1 percent. That is, again there was a large surge in 1968–1969 that persisted until 1971.

II.F. Conclusion

Looking only at the expectations from the Federal Reserve staff and the cross-sectional mean of professional forecasts or households, the inflation expectations anchor seemed to be firm until 1971, with the ever-rising inflation being treated as a sequence of unfortunate temporary deviations. Only after the end of Bretton Woods there is a sustained drift upward.

This section noted that looking at other, arguably richer, data paints a different picture. Already between 1967 and 1970, the anchor went adrift as revealed by market prices for gold futures, public concerns and media reports, and, perhaps more convincingly, by the shift in the cross-sectional distribution of household expectations.

After mid-1970, household expected inflation fell even as professionals and markets were rising, perhaps because of the wage and price controls. This effect was temporary, and by 1973 the inflation expectations anchor was definitely lost.

III. A Measure of Fundamental Expected Inflation

In almost every dynamic macroeconomic model, both actual inflation and inflation expectations are endogenous objects that are closely related. Under the extreme assumption of perfect foresight, they coincide, while with full-information rational expectations, the difference between the two is serially uncorrelated and has a mean of zero. Therefore, a traditional approach to measure the expected inflation is to measure the actual inflation. Then,

given a goal of foreseeing future inflation movements, since expected inflation is their anchor, economists have focused on current inflation as a predictor, supplemented with its lags. But the Lucas critique (Blanchard 1984) dictates that the relation between expected inflation and lags of inflation can change and that it is unreliable, especially when inflation is about to drift away.

This paper has taken a different approach, measuring inflation expectations directly and investigating whether they provide signals about future inflation. I found that they do, but only if one uses the whole cross-sectional distribution of survey expectations from both households and professionals complemented with market-price expectations. It would be useful to combine these into a single estimate of expected inflation. This requires having a model of expectations, and there is a myriad of them, each with its own virtues and flaws. However, almost all of them share one feature: expectations ultimately converge to the rational expectations fundamental. Whether agents use past data to learn or have incomplete signals, limited attention, or heuristic biases, their expectations will differ from rational expectations but gradually converge with them.[16] If an economic fundamental, like a policy regime or a shock, makes the rational expectation change, then in almost all modern models expected inflation will change in the same direction as well. The rational expectation serves as a fundamental for expected inflation.

III.A. A Model as a Measurement Tool

This section uses a parsimonious model following Reis (2021) to combine the different expectations data and estimate this fundamental. The goal is not to test theories or distinguish between models of expectations but to combine their key ingredients to extract as much signal as possible from the data. The model has three equations for three measures of expected inflation:

$$(1)\quad v_t^h = \pi_t^* + c_t^h + \theta_t\left(e_t^h + \pi_t^e - \pi_t^*\right) \text{ with } c_t^h \sim E\left(\lambda_t\right),\ e_t^h \left| \pi_t^e \sim N\left(0, \sigma_t^2\right)\right.,$$

$$(2)\quad q_t = \frac{\int y_t\left(\pi_t^e\right) g_t\left(F_t^{-1}\left(\omega_t\right)\right) f_t\left(F_t^{-1}\left(\omega_t\right)\right) d\pi_t^e}{\int g_t\left(F_t^{-1}\left(\omega\right)\right) f_t\left(F_t^{-1}\left(\omega_t\right)\right) d\pi_t^e} \text{ with } \omega_t \sim B\left(\beta\right), \text{ and}$$

$$(3)\quad E_t^h = \mathrm{E}_t\left(\pi_t \left| v_t^{median}, q_t\right.\right).$$

16. Most models of expectations eventually converge to the rational expectations equilibrium, much like most business cycle models eventually converge to the Ramsey-Solow steady state.

The fundamental rational expectation of inflation is π_t^e. Each equation reports how the expectations of households v_t^h, market prices q_t, and professionals E_t^h relate to it.

The first equation implies that the expectations of inflation v^h reported by a household h in a survey follows an exponentially modified Gaussian (EMG) distribution $F_t(.)$, which is the sum of three parts: The first part is just a prior mean π_t^* for what fundamental expected inflation π_t^e might be. The second is a bias c_t^h drawn from an exponential distribution. It is important to match the fact that survey expected inflation is persistently above actual inflation, perhaps as a long scar from the high inflation of the 1970s and 1980s. The third is an individual-specific signal of the fundamental, with noise e_t^h, drawn from a normal distribution. The relevant parameters are the reaction of expectations to signals θ_t, the informativeness of the signal affecting the dispersion of expectations σ_t^2, and the average bias $1/\lambda_t$. Reis (2021) shows that this model captures the main properties from the large literature on imperfect information, overreaction, experience learning, and sticky information.

The second equation states that the market price expected inflation q_t is equal to the expected risk-adjusted inflation with risk adjustment $y_t(\pi_t^e)$. That expectation is formed by traders who start from a prior which is the EMG $F_t(.)$ (traders are people after all) but which is updated following Bayes's rule after observing the market price itself, as written out in equation (2). The price is contaminated by a noise ω drawn from a symmetric beta distribution with parameter β. The density $g_t(.)$ is the endogenous distribution that results from markets having to clear given the noise and fundamental. Reis (2021) shows that this model proxies for many models of imperfect information in financial markets that are in the spirit of Grossman and Stiglitz (1980).

The third equation states that the professionals in the surveys are similar to traders of inflation risk. They use Bayes's rule to combine the information that households have with market prices, but what they report in a survey is neither risk-adjusted nor identifies who is the marginal player in the markets. Rather, from the small samples in these surveys we can only see the objective beliefs of the median trader.

III.B. The Information in Different Data and the Model to Filter It

Individually, each different source of expectations data carries useful information. The household expectations are sluggish, less informed, and subject to biases, but we have more cross-sectional data with which to estimate

the three moments of their cross-sectional distribution and to separate the fundamental from bias and noise. The market price is erratic and polluted by financial noise, but it is fast to react to fundamentals. Surveys of professionals or traders have small samples, so only a first cross-sectional moment can be calculated. But, combined with market prices, they can separate the fundamental from the market noise. While each series has virtues and flaws, it is by combining all of these pieces of data that the model is able to uncover the fundamental expected inflation.

With observations of the first three moments of the cross-sectional distribution of survey inflation expectations, together with a market price for expected inflation, and a first moment from a survey of professionals or traders, one has five numbers at each date. With these, and the three equations above, one can back out the reaction, dispersion, and bias coefficients θ_t, σ_t^2, λ_t, as well as financial market noise ω_t and fundamental expected inflation π_t^e. The model gives a way to filter these five pieces of information and recover the anchor for expectations that underlies all of them.

If the market prices' expected inflation rises, but there is no change in survey expectations, the model will attribute most of it to financial market noise and not change the estimate of fundamental expected inflation. At the other extreme, if the cross-sectional mean of survey expectations of households and professionals both moved in one direction, the model would update its π_t^e in the same direction. In between, and more relevant, if the cross-sectional household survey mean barely moves, but the cross-sectional survey standard deviation and skewness of households moves together with the market survey and price-implied expectations in a consistent direction, then the model infers that the fundamental has changed. Some people have started updating, and if they have done so in the same direction as traders, then the fundamental must have changed.

To take the model to the 1966–1971 annual sample I use the three moments from the Michigan quantitative survey, the median expectation from the subset of professionals in the Livingston survey that work in the finance industry, and the annual expected inflation implied by the Zurich gold futures. From 1972 onward, there is no clear market price to use to implement the model. Note that all inputs are series of expectations; nowhere does actual inflation enter the calculations.[17]

17. Carvalho and others (forthcoming) provide an alternative measure of the inflation expectations anchor using actual inflation data.

For the prior inflation π_t^*, I use the previous year's fundamental inflation estimate π_{t-1}^e. This way, the econometrician's prior is to find no change in the fundamental. Finally, the parameter β measures how noisy financial prices are. Following Reis (2021), I set it equal to 2, which allows for a significant amount of noise in financial markets.

There are three important limitations in using this model with the 1967–1973 data. First, the data on market prices come from the Zurich gold market, but the data on surveyed market participants come from professional economists in the United States. These are not as close to each other as would be desirable. Second, there are no data with which to calibrate changes in the inflation risk premium, so I assume it is constant (and since the model only identifies changes, I set it to zero: $y(\pi^e) = \pi^e$). Third, as discussed in Reis (2021), this is a model of long-horizon expectations. For short-horizon expectations, the different structural models of expectations in the literature are too diverse to be captured in a parsimonious way. Yet, there are no long-horizon expectations data. So, I treat the data on one-year expectations as proxies for longer-horizon expectations.

III.C. The Estimates of Fundamental Expected Inflation

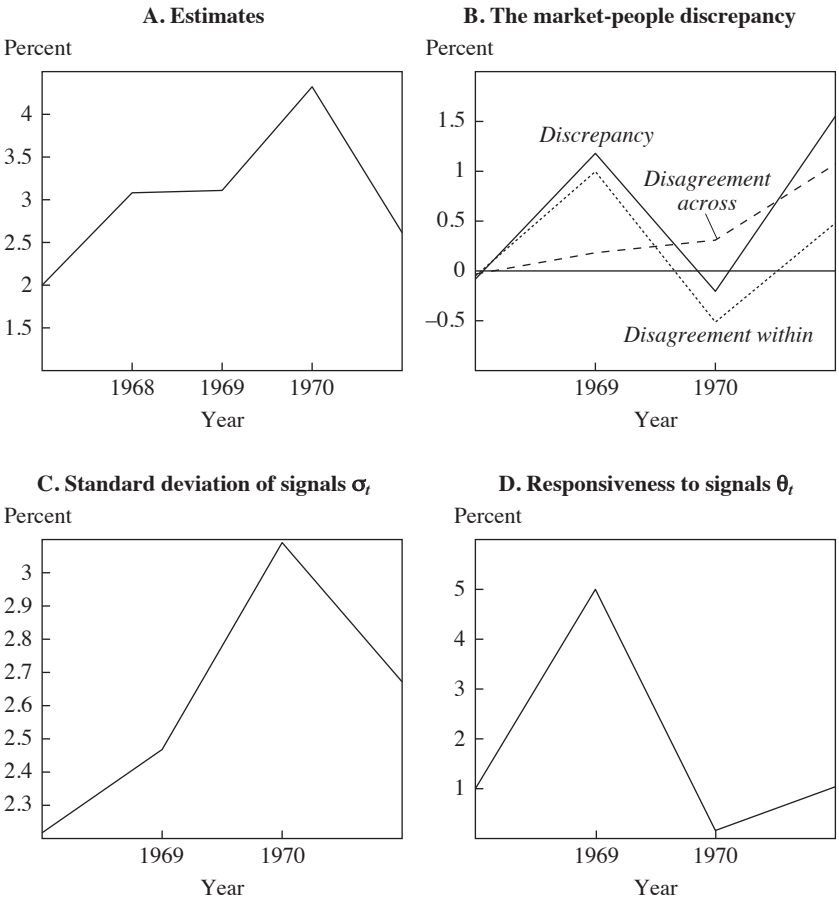
Figure 10 plots the estimates of the model. Panel A shows the fundamental expected inflation. It rose between 1967 and 1970 by 2.3 percentage points. During this time, actual inflation rose by 3.1 percentage points.[18] The bond yields or the Greenbook forecasts that policymakers looked at during that time would have given a false sense of stability. Instead, the rising market prices combined with the large increase in standard deviation and fall in skewness across households reveal that fundamental expected inflation was changing significantly. The estimates also confirm the temporary respite from this increase in 1971.

To understand what is going on, panel B plots the difference between the market expectation and the household expectation and decomposes it into the difference within traders (marginal versus median) and across people (professionals versus households). At first, market expected inflation rose, while surveys of professionals showed little change. The model interprets this disagreement within as partly due to a financial shock. However,

18. The initial level of expected inflation is not identified. It is pinned down by the choice of prior π^* in the initial period, so the appropriate comparison between actual and expected is their predicted changes. In the figure I set initial expectation in 1967 to 2.0 percent, even though actual inflation was 3.2 percent.

15985-05a-Reis-4thPgs.indd 336 5/12/22 12:07 PM

Figure 10. Fundamental Expected Inflation



Source: Author's calculations.

because of the increase in variance and fall in skewness that was shown in figure 6, panel A, the model puts a significant weight on this instead being a result of expected inflation rising. Therefore, the estimate of expected inflation is updated upward. The following year, this suspicion is confirmed, as the median expected inflation rises across surveys converging to the market expectation. In 1970, the rise in survey expectations combined with the movements in disagreement confirm that the fundamental has risen, so the model adjusts its estimate upward again, even though market prices have fallen.

Panels C and D show the time series of two important model parameters, σ_t, which measures the dispersion of the idiosyncratic signals, and θ_t,which measures by how much an individual will respond to his/her individual signal. If people were pure Bayesians, then θ_t would be proportional to the inverse of σ_t, but the model allowed this to not be the case following theories of overconfidence, overreaction, and diagnostic expectations. The estimates strongly support this. Throughout the period, the estimated standard deviation rises. Agents are confused and disagree on where inflation is heading. But at first, they are also more responsive to their individual signals, as if they were paying more attention to the inflation path in 1969. Only after, in 1970, do agents return to their usual responsiveness. Finally, 1971 continues to stand out as a time where there was a reversal in these drifts.

IV. Measuring the Expected Inflation Anchor in Other Episodes

Measures of expected inflation are sufficiently noisy that using them to forecast inflation during tranquil times will likely bring little benefit. The previous two sections showed that when there was a regime change in the United States and inflation persistently drifted, then the expectations data were useful to diagnose this shift. This section investigates whether the same is true in other episodes of regime change, when inflation moved significantly and persistently. Each episode provides a different test of our ability to measure the expected inflation anchor.

Searching through the history of inflation over the last fifteen years, there were only a handful of countries that had both good survey data and data on market-price expectations of inflation, as well as inflation above their central banks' targets for a few years. Three episodes stand out. Brazil between 2011 and 2016 provides the clearest counterpart to the US inflation of the 1970s. Turkey has also had a drifting inflation since 2018, but because it is more recent, the jury is still out in terms of how persistent this will be. Looking at the expectations data today in 2021 gives an example of how to try to measure the anchor in almost real time. Finally, South Africa between 2010 and 2016 provides a placebo, or a possible example of whether the expectations data can deliver false positives. While inflation was high during those years, this turned out to be temporary, and the rate of price changes was back on target afterward.

After these three cases, this section returns to US history to look at what the expectations data showed in the early 1980s, when inflation came down

sharply. Finally, it concludes with a real-time look at expectations data after the 2020 pandemic.

IV.A. Brazil, 2011–2016

After two decades of struggling against high, and sometimes hyper, inflation, Brazil managed to reform its monetary regime in 1994 and to bring inflation down (Ayres and others 2019). In 1999, the Banco Central do Brasil (BCB) adopted an inflation targeting regime with a target for the CPI and wide upper and lower bounds that are periodically updated. Between 2002 and 2003 the regime came under a trial by fire, as the combination of a domestic drought, contagion from the financial crisis in Argentina, and a contested presidential election cycle led to a sharp increase in sovereign spreads, a depreciation of the currency, and inflation spiking to 17 percent in 2003, well outside the target range. Aggressively tight monetary policy followed, and inflation came down sharply back to target.

ACTUAL INFLATION In August 2010, the inflation rate—year-on-year changes in the Extended National Consumer Price Index (IPCA)—was 4.5 percent, exactly its target. The BCB had just gone through a tightening cycle, but inflation was modestly rising, perhaps due to large public deficits and increases in oil and commodity prices. A new president came into power in January 2011, and in the August policy meeting, according to press accounts, the BCB unexpectedly and significantly eased policy. The main policy interest rate (Selic) was reduced from 12.5 percent to 7.25 percent continuously over the next fourteen months, this in spite of GDP in 2010 having risen by a remarkable 7.5 percent, and inflation in July 2011 having been 6.9 percent, above the upper bound of the target of 6.5 percent.

Panel A of figure 11 shows the evolution of Brazilian inflation. Between 2011 and the end of 2015, inflation was always close to the upper bound, peaking in January 2016 at 10.7 percent. Public deficits were large and growing, and the government resorted to hidden fiscal expansions by persuading the main oil producer to sell domestically below market prices and having the state-owned development bank (BNDES) and mortgage lender (Caixa) run persistently large losses that amounted to a quasi-fiscal deficit. Monetary policy was timid in its response to inflation rising, only starting to gradually raise the interest rate after May 2013. This led to concerns about fiscal dominance of inflation and lack of independence of the BCB, especially as the central bank reports and policy minutes were persistently optimistic about inflation, fiscal consolidation, and the public credit subsidies. Moreover, the government kept the administered prices of gasoline and diesel low between 2011 and 2014 and cut energy tariffs and sales

Figure 11. Brazil's Drifting Expected Inflation Anchor: 2011–2016

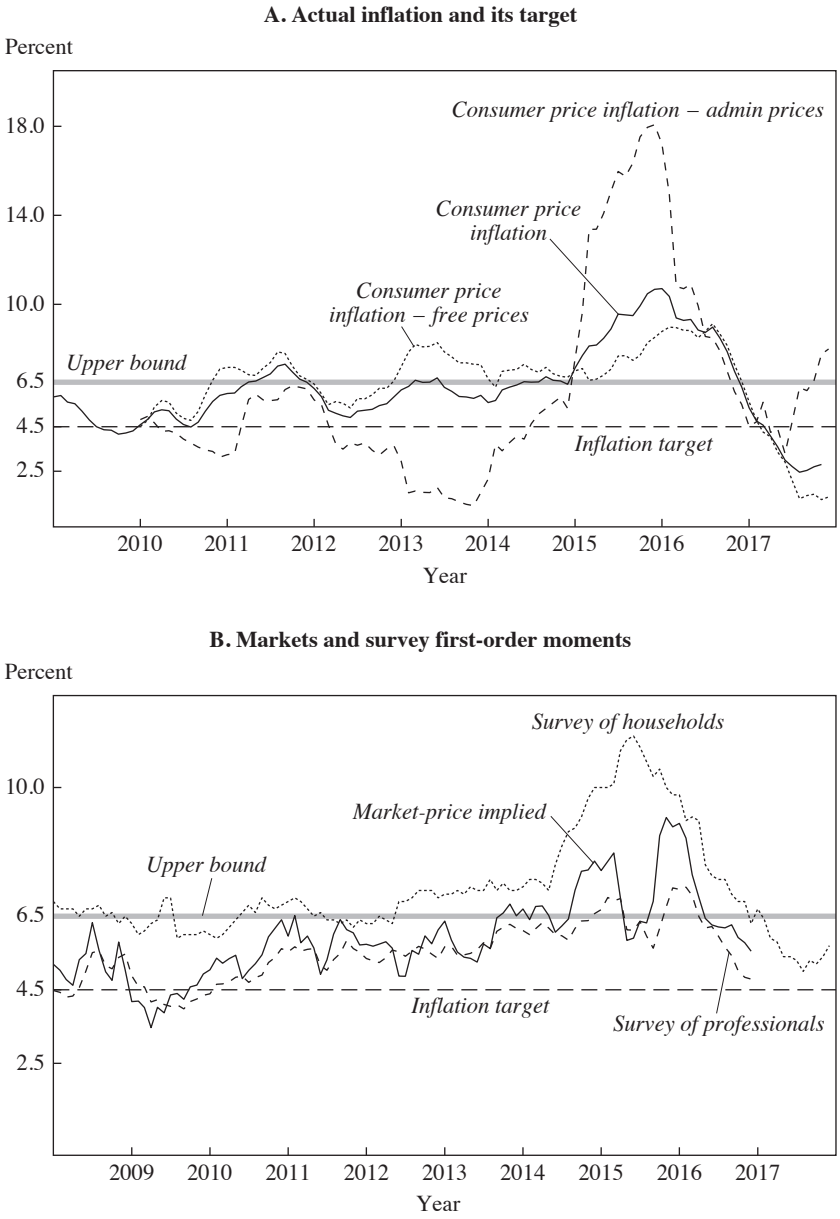
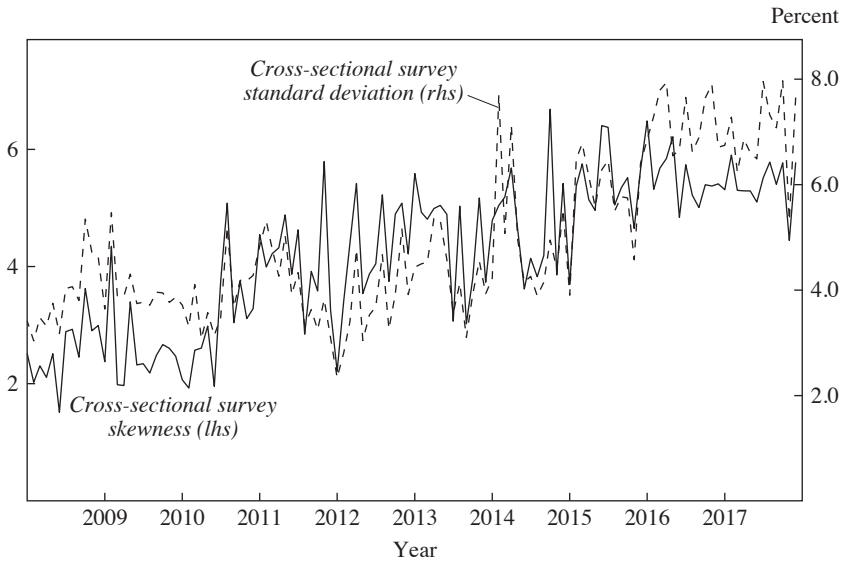
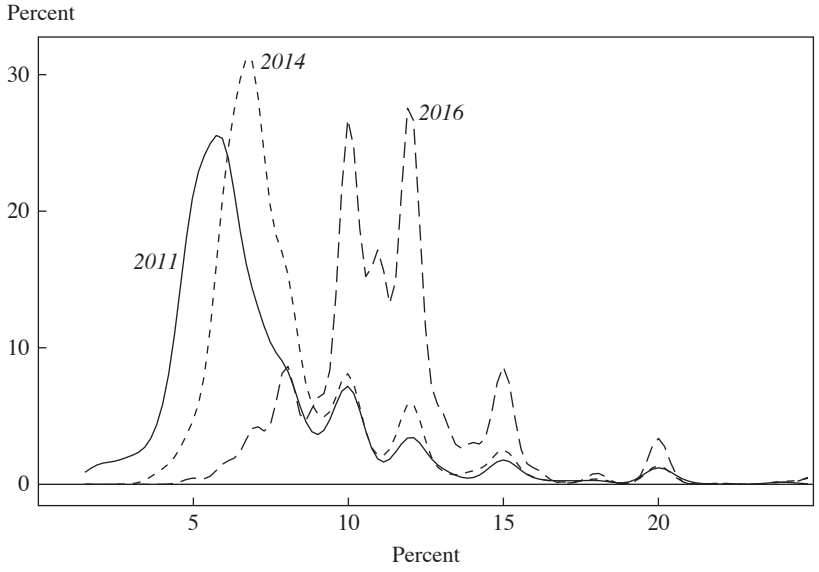


Figure 11. Brazil's Drifting Expected Inflation Anchor: 2011–2016 (*Continued*)

C. Cross-sectional disagreement of households



D. Cross-sectional distribution of households



Sources: Banco Central do Brasil; Focus Survey by Banco Central do Brasil; FGV IBRE (https://portalibre.fgv.br/en/fgv-dados); Val and Araujo (2019).

taxes for some commodities, all directly lowering measured consumer prices. This created a large gap between prices that were set in markets and those that were set by the government, which contributed to rising public deficits (Bonomo 2018).

After the reelection of President Dilma Rousseff in October of 2014, administered prices were adjusted back in line, which led to a large jump of overall inflation by 3.3 percentage points over the next twelve months. Accumulating the change in the price level between August 2010 and August 2016, the annualized rate of inflation was 7.3 percent, well above the 6.5 percent target. After a controversial impeachment process, Rousseff left office at the end of August 2016. The new government replaced the members of the monetary policy committee, monetary and fiscal policy were tightened, and by March 2017 inflation was back at the 4.5 percent target. Some good luck followed in the form of strong agricultural production lowering food prices, and inflation was always below 3 percent between mid-2017 and mid-2018.

EXPECTED INFLATION Panel B in figure 11 shows series for inflation expectations in Brazil during this time. The professional forecasters series corresponds to the median answer in the Focus survey ran by the BCB.[19] It is an unusual survey because it includes as many as 140 respondents drawn from financial market participants, and it provides incentives by giving out a prize every year to the most accurate forecaster, which is then widely reported in the financial media. The household expectations correspond to the survey of consumer sentiment that includes approximately two thousand consumers and asks a precise numerical question about inflation over the next twelve months. The series corresponds to the mean over the sample that excludes the top and bottom quartile of answers and is reported monthly by the Brazilian Institute of Economics of the Getulio Vargas Foundation (FGV-IBRE). Finally, the market price expected inflation refers to the breakeven rate—the difference between the rates in inflation-indexed and nominal bonds—over the next year, corrected for seasonal factors and indexation lags.[20]

Both the survey of professionals and the market prices track actual inflation quite well. This is perhaps not too surprising, given the strong incentives

19. Banco Central do Brasil, "Market Expectations—Focus Survey," https://www.bcb.gov.br/en/monetarypolicy/marketexpectations.

20. For more on the Focus survey, see Alves, Areosa, and Carvalho (2020); on the FGV-IBRE series, see Gaglianone, Issler, and Matos (2017); and on the breakeven series, see Val and Araujo (2019).

to do so. Insofar as inflation was being kept artificially low through the administrative prices, this is less informative about the anchor. The first moment of the household survey starts clearly drifting up from the start of 2013 onward. It strongly suggests that the anchor was drifting away.

Panel C shows the second and third cross-sectional moments from the household survey.[21] This shows a clear increase in both series since the middle of 2011. The starting positive skew in the Brazilian household expectations is much lower than its counterpart in the United States. This is consistent with a smaller inflation bias and higher rate of information updates by Brazilian citizens given the country's history.[22] But during this period, the loss of the anchor appears, as more households move to the right of the distribution. This both raises the variance and increases the positive skew.

Panel D confirms this by plotting the distributions at the start of the year in 2011, the middle of 2014, and the start of the year in 2016. The increase in variance and skewness clearly arises from a subset of households breaking off to the right of the median. They expected higher and higher inflation as months went by.

LESSONS The Brazilian experience in 2011–2016 adds to the lessons from the US experience of 1967–1974. A similar combination of timid monetary policy, fiscal dominance, and persistent belief that inflation is only temporarily high were the harbingers of rising inflation. Moreover, the government's direct interference with the CPI via administrative prices lowered professional and market price expectations but appear to have only delayed the movements in household expectations. Like in the United States, such policies seemed to be effective only temporarily.

As in the United States, measures of disagreement in the cross-sectional survey provide a good image of the loss of the inflation anchor, but one should not expect them to be captured by a naïve time-series correlation between standard deviation and skewness. In Brazil they moved together, while in the United States initially skewness and variance moved in opposite directions. The reason is revealed by the cross-sectional distribution

21. I calculate this by using the micro data directly, excluding only the top and bottom 1 percent of answers to deal with outliers. Results cutting the top and bottom 5 percent are very similar. Also, the mean calculated this way is not too different from the mean calculated with the official procedure of excluding the bottom and top 25 percent (which would be too aggressive to calculate higher-order moments).

22. Fitting the EMG distribution from section III, the λ parameter is much higher for Brazil than for the United States.

plots. The initial drift of the anchor in the United States came with a thinning out of the left tail of the distribution: those who expected very low inflation joined the median. In Brazil, the median instead moved to the right and the right tail thickened. A naïve correlation of these two measures of disagreement will not be particularly informative.

IV.B. Turkey, 2018–Present

Turkey had a brush with very high inflation during the 1990s associated with political instability and the monetary financing of deficits. After an IMF program and a law giving independence to the Central Bank of the Republic of Turkey (CBRT) in 2001, a sharp disinflation program lowered it from 54 percent in 2001 to 8 percent in 2005 with little output costs. In 2006, an inflation targeting regime was adopted, with a target around 5 percent, although with some annual changes. Actual inflation was always above target but steadily so, averaging 8 percent between 2006 and 2017.

ACTUAL INFLATION In the three years between the start of 2018 and the end of 2020, inflation shot up to an average of 15 percent, three times the inflation target (Kara 2021). The precipitating event behind the rise in inflation seems to have been the reelection of Recep Erdoğan in June 2018. This was a period of political instability, following a failed coup in July 2016, a constitutional referendum in April 2017, and the premature 2018 election that should have taken place in November 2019. As the president consolidated his power, the economy was faltering, and he started commenting on inflation and interfering with the CBRT's independence. In May 2018, in a campaign speech in London, he expressed desire to take greater control of the economy and defended lower interest rates in order to control inflation. In June, he announced he would investigate Moody's after a downgrade of the country's debt, and the lira depreciated strongly against the US dollar. In September, Erdoğan repeated his theory that lower interest rates are desirable and that inflation was not a monetary phenomenon, sharply criticizing the CBRT.

On the fiscal side, defense spending has increased significantly since 2015, associated with the wars in Syria and Iraq. The early elections of 2018 came with several new spending measures in 2017–2018: 750,000 contract workers in the public sector became permanent employees, a minimum wage subsidy estimated to cost 0.2 percent of GDP was extended together with a monthly subsidy for new hires, and a bonus was offered to all pensioners at two religious holidays costing 0.7 percent of GDP (IMF 2018; OECD 2018). As for monetary policy, right after the election in July 2018, the CBRT surprisingly did not raise interest rates, as had been

15985-05a-Reis-4thPgs.indd 344 5/12/22 12:07 PM

widely expected since inflation was at a fourteen-year high. However, in September, it sharply increased from 17.75 percent to 24 percent, right after the criticisms from Erdoğan.

EXPECTED INFLATION Panel A of figure 12 also shows monthly inflation expectations from *t* to one year ahead. The market price series is a measure of breakeven inflation rates made available by the CBRT. Throughout, they have been very significantly underpredicting inflation.

The second series is the median from a monthly survey of professionals that is dominated by financial market participants who are asked to put their expected inflation in bins. This survey is also collected by the CBRT, starting in 2001, and has sixty to seventy respondents. The figure also shows the median response from the Business Tendency Survey, also conducted by the CBRT since 1987 and that every month surveys approximately five hundred firms on business confidence, including a question about Producer Price Index inflation.[23]

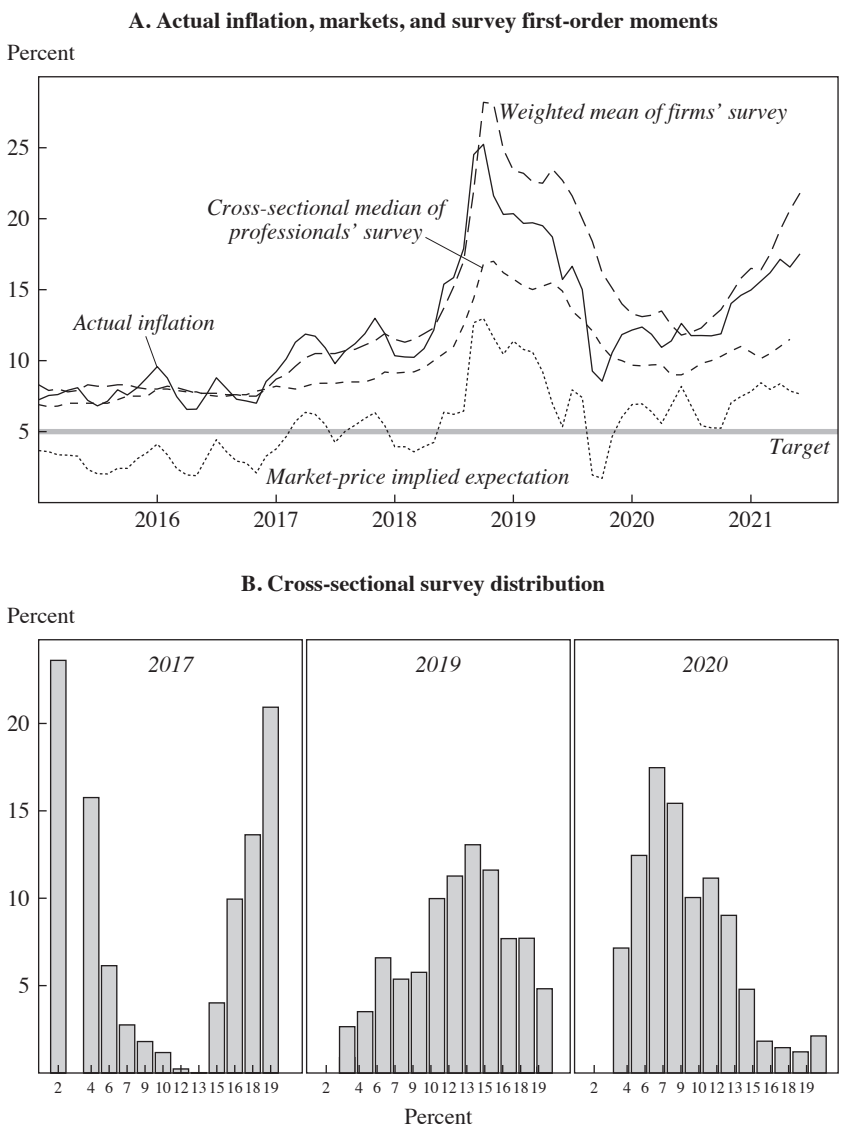
All three measures of expected inflation move closely together. All of them missed the 2018 burst in inflation back in 2017, all of them expected some of that inflation to revert back, and all of them expected some of it to persist into the future

The micro data from the survey of firms are not available. An imperfect proxy is the survey of professionals. This is imperfect for three reasons. First, these are professionals, and at least for the United States, their expectations tend to be closer to normal and less informative beyond the cross-sectional mean. Second, the number of respondents is not large enough to accurately calculate a third-order cross-sectional moment. And third, the survey's top bin is 20 percent or more, which given the rise in inflation will truncate the right tail.

Subject to all these caveats, already by the end of 2017, the standard deviation almost quadrupled, while the skewness went from being negative at −1 percent to positive at 0.25 percent. Panel B of figure 12 shows the distributions in December 2017, January 2019, and June 2021. In 2017, the uncertainty is evident, with a bimodal distribution and more than half of the respondents expecting inflation to exceed 17 percent. The events of 2018 removed some of the disagreement by consolidating a view that inflation would be well above the target. By 2021, more mass has moved rightward, and the inflation anchor is definitely lost.

23. CBRT, "Tendency Surveys," https://www.tcmb.gov.tr/wps/wcm/connect/EN/TCMB+EN/Main+Menu/Statistics/Tendency+Surveys.

Figure 12. Turkey's Drifting Expected Inflation Anchor: 2018–2021



Source: Central Bank of the Republic of Turkey, Electronic Data Delivery System (https://evds2.tcmb.gov.tr/index.php).

LESSONS The Turkish experience of a lost anchor leads to two additional lessons. First, even close to real time, and when inflation is bouncing up and down, like it did in Turkey in 2018 and 2019, the expectations data can paint a clear picture of a lost inflation anchor. Second, in countries where arguably the anchor was not firm in the seabed to start with, shifts in the cross-sectional distribution can be large and fast. The loss of the inflation anchor can come fast and need not be gradually building up like it did in the United States in the late 1960s.

IV.C. South Africa, 2010–2016

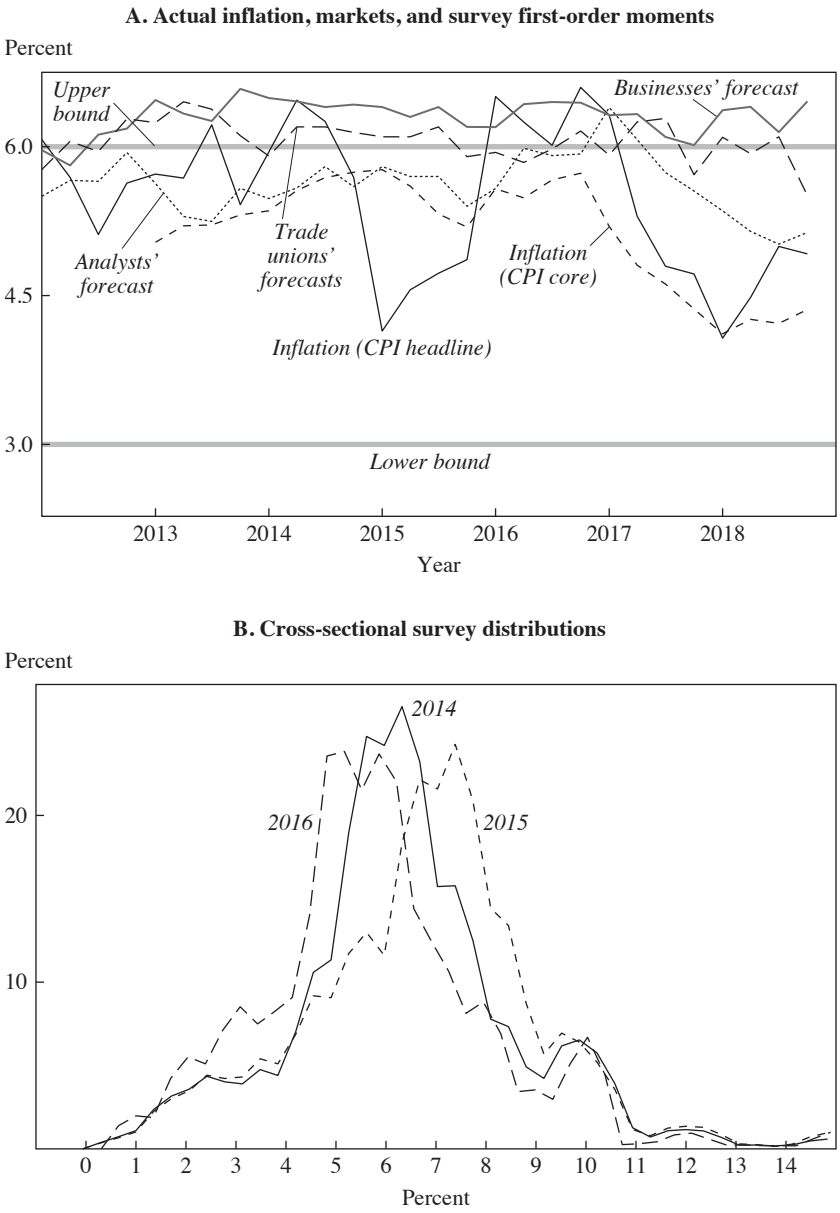
The South African Reserve Bank (SARB) adopted inflation targeting in 2000, with a target range of CPI inflation between 3 percent and 6 percent and no stated midpoint. The first few years of the new regime were rocky, with oscillations in the exchange rate and reversals of policy, but after 2005 transparency increased, so that after one decade of inflation targeting, outcomes were on average solidly within the range. The global financial crisis of 2008–2010 brought a recession, affecting the country mainly through its international trade linkages to Europe.

ACTUAL INFLATION Between 2010 and 2016, the SARB had to face an unusual, difficult sequence of supply shocks. Inflation averaged only slightly less than the 6 percent upper bound of its target as a result, as shown in panel A of figure 13. First, the rise in global oil and commodity prices in 2010–2012 pushed inflation toward the upper bound of the range. At the same time, the recession in Europe due to its sovereign debt crisis implied that the South African economy was slow to recover. The SARB kept interest rates steady so as to not prevent that recovery.

Second, as inflation was falling in the first half of 2012, with the economy recovering and oil prices stabilizing, inflation was hit by a different shock throughout 2013 and 2014. The state-owned electricity provider ran perennial deficits and needed large investments to solve the persistent outages. The increase in costs led the regulators to approve a large increase in prices, well above the inflation targeting range. Headline inflation continued to exceed core inflation, but monetary policy held steady for a while until finally raising the repo interest rate in 2014, for the first time in six years. By 2015 these effects were over and inflation fell sharply to the middle of the target range, leading the SARB to loosen monetary policy.

A third shock arrived soon though. In 2016, an El Niño storm together with a drought led to a large decline in soil productivity, which raised food prices that affect a large component of the consumption baskets. Again, headline inflation moved to the upper range of the target, even as core inflation

Figure 13. South Africa's Unlucky Run: 2010–2016



Sources: Statistics South Africa (http://www.statssa.gov.za); Bureau for Economic Research, "Inflation Expectations (2021)" (https://www.ber.ac.za/BER%20Documents/Inflation-Expectations-History-(Excel)/?doctypeid=1099&year=2021); Reid, Du Plessis, and Siklos (2021).

was well within it, and again the SARB held monetary policy steady. After 2017, no more shocks arrived, and since then inflation has stayed within the target range.[24]

EXPECTED INFLATION Panel A of figure 13 shows the median expected inflation from three separate samples: financial market analysts, nonfinancial firms, and trade unions. The survey is conducted by the Bureau for Economic Research (BER 2021) for the SARB on a South African panel quarterly since 2000. All three behaved similarly during this period, persistently forecasting inflation to be around 6 percent. On the one hand, these forecasts by professionals proved to be reasonably accurate ex post, and they are remarkably stable in light of the volatility in actual inflation. On the other hand, they missed the fall in inflation in 2015 and were slow to catch up to the lower inflation from 2017 onward.

Panel B reproduces the cross-sectional distribution among households at three successive Octobers between 2014 and 2016, calculated by Reid, Du Plessis, and Siklos (2021). These come from a remarkable survey of between 2,000 and 2,500 individuals in urban and rural environments at a quarterly frequency, conducted by AC Nielsen under contract with the SARB and the BER. As inflation moves up and down, the distributions shift right and left. However, note that disagreement, measured by either second or third moments, does not change much.

LESSONS The survey data throughout the 2011–2016 period seemed consistent with a stable anchor. Shocks hit the economy, the central bank justifiably kept monetary policy steady, letting inflation rise, and both outcomes and expectations reflected this with higher expected inflation. Yet, there was no permanent rise of either actual or expected inflation, unlike what we saw in the United States, Brazil, and Turkey. Disagreement did not increase during this period. Contrary to the United States in the 1970s, inflation did not drift up as the expected inflation anchor remained steady. In the South African case, the price controls worked in the opposite direction of what they did in the United States and Brazil, yet their effects were qualitatively similar: significant but temporary.

IV.D. The United States in the 1980s: Dropping the Anchor

Between 1979 and 1983, under chair Paul Volcker, the Federal Reserve adopted highly restrictive monetary policy and inflation fell significantly. If the main episode studied in this paper is the loss of an inflation anchor,

24. For a description of the SARB's reaction to the successive shocks, see Kabundi, Schaling, and Some (2015) or Mminele (2019).

this next episode is the dropping of a new anchor, which persists today. It adds a reversal situation and again tries to measure the anchor.

Panel A in figure 14 shows that the survey of households, which now corresponds to the standard Michigan quantitative series, was quick to catch on. The decline was swift, keeping up with inflation. Professional forecasters were slower (or perhaps more skeptical). Digging deeper into the household cross-sectional distribution, panel B shows that as some households started expecting lower inflation, this increased the standard deviation, while it lowered the positive skew. As the remaining households gradually caught up, the median fell, the standard deviation started declining after reaching a peak, and the skew started rising. Altogether, this behavior is consistent with the model described in section III, where people who are inattentive, overconfident, and use sticky information in updating their biases would react in this gradual way to a change in policy regime. As in section III, it shows that looking at the distribution of expectations, from first to third moment, can provide some signals of where the expected inflation anchor is and where it is going.

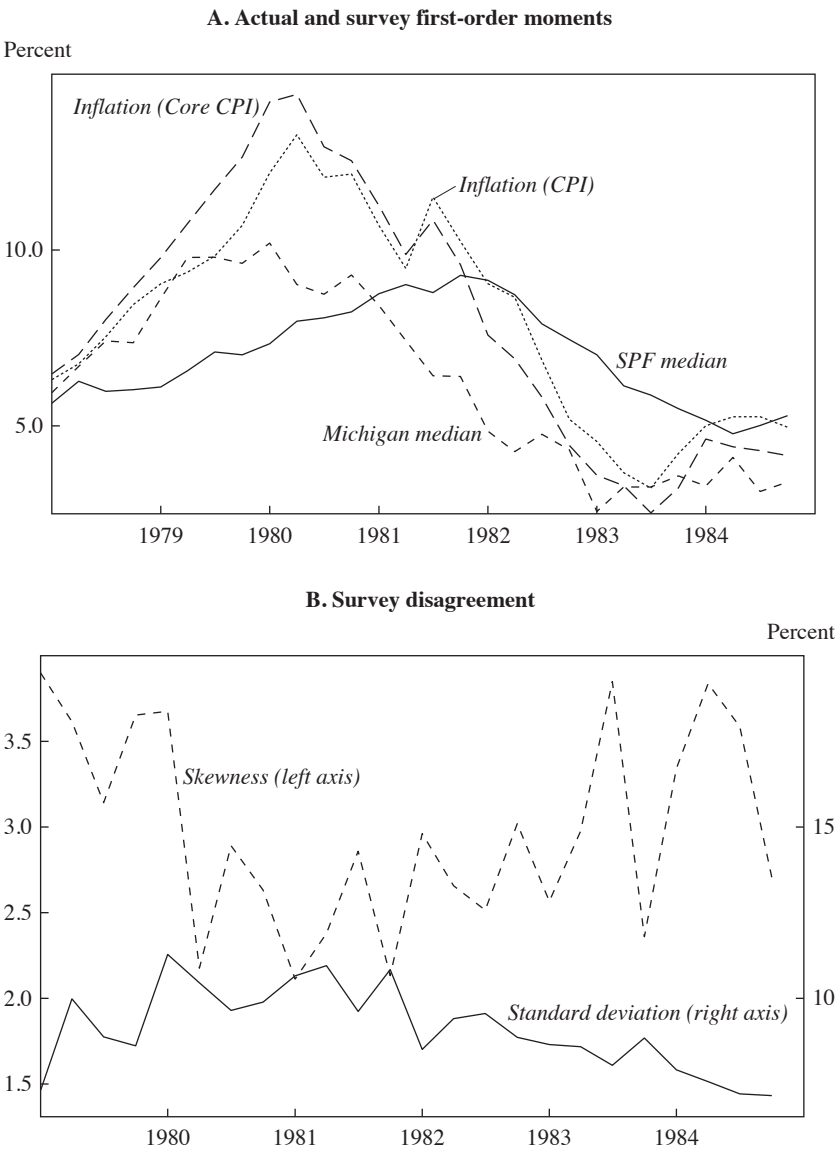
IV.E. United States 2020–2021: Where Is It Heading?

The final application of the ideas in this paper is more tentative. The pandemic recession of 2020 and the swift recovery in 2020–2021 interrupted three decades where the expected inflation anchor was steadily at 2 percent and actual inflation had only small transitory movements near its anchor. As panel A in figure 15 shows, inflation fell sharply with the lockdown in the first half of the year and rebounded very strongly in the first half of 2021 reaching levels that had not been seen for decades. Today, many ask the question that this paper asked at the start: Where is inflation heading?

The economic disruptions associated with the pandemic have many peculiar features that make this a hard question to answer. Still, we can look at the expectations data to see if the same patterns are present as in the other episodes of a lost anchor. Panel B in figure 15 shows that professional forecasters have raised their expectation only slightly. If they were the single measure of the anchor, one would confidently conclude that inflation will soon quickly revert back to this anchor. Market price expected inflation instead has increased significantly in 2021, albeit only after having fallen significantly in 2019 and 2020. These sharp movements in market expectations might shake that confidence, but they could be dismissed as another illustration of the market noise ω in the model.

The household survey is more worrying. Over a period of only six months, both the long-horizon and the one-year-ahead inflation expectations

Figure 14. Dropping the Anchor: The United States in the 1980s



Sources: Mankiw, Reis, and Wolfers (2004), from Surveys of Consumers, University of Michigan, and FRED.

Figure 15. The Expected Inflation Anchor through the Pandemic

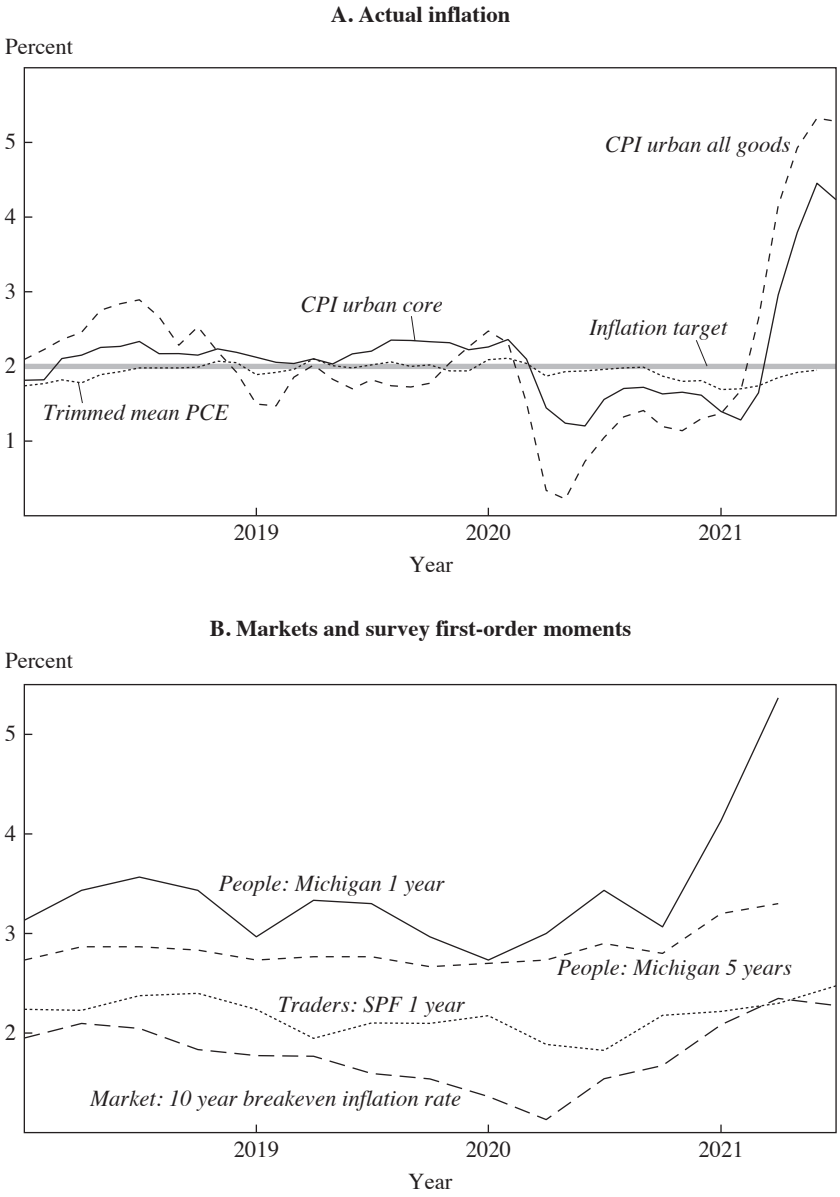
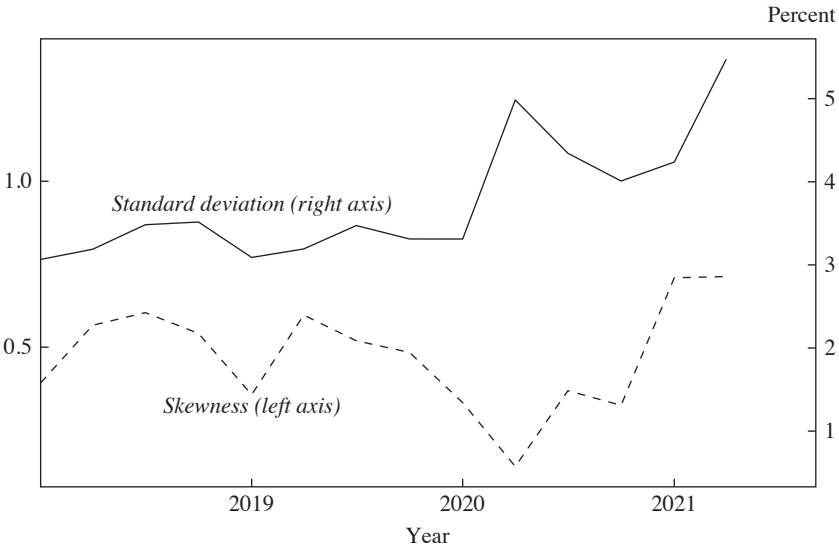
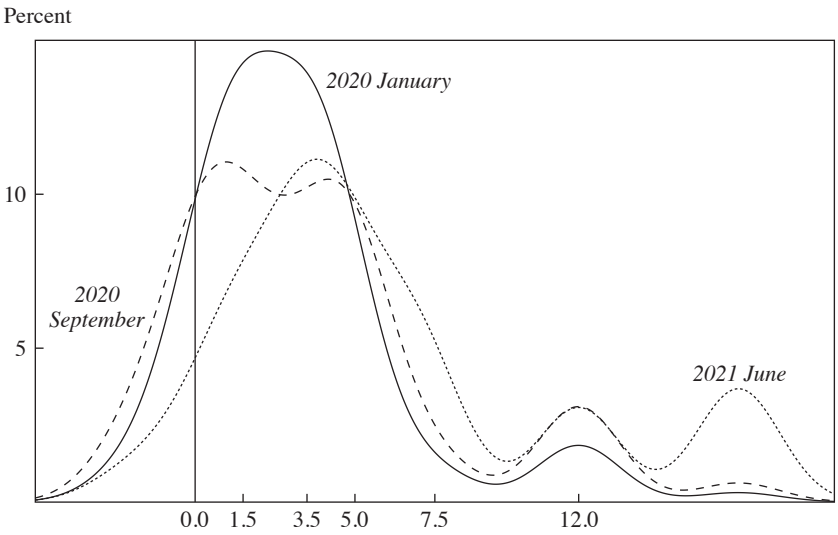


Figure 15. The Expected Inflation Anchor through the Pandemic (*Continued*)

C. Cross-sectional disagreement of households



D. Cross-sectional distribution of households



Sources: FRED; author's calculations from Surveys of Consumers, University of Michigan.

have jumped up faster than almost ever before in the post-1980s sample. Given its usual sluggish adjustments, this unusual quick jump is hard to interpret. Candia, Coibion, and Gorodnichenko (2021) report a very high discrepancy between the expectations of households and professionals across different surveys and for other countries as well.

Again, looking at higher-order moments and, more generally, at the whole distribution provides some guidance. Between the start of 2020 and the second half of the year, mass moved away from the peak of the bell-shaped distribution and toward the right tail. That is, some households started believing that inflation might be very high. As a result, both the standard deviation and the skewness increased rapidly. In the first half of 2021, it is instead the left tail that has been hollowed out. Few US households today believe that inflation will be below target in the near future, while many expect that an inflation disaster could happen. As Knotek and others (2020) document, households have been particularly attentive and responsive to news about inflation during this time period.

Perhaps, like in the early 1980s, households are ahead of the curve in detecting a change in the policy regime. Or perhaps through the overconfidence discussed in the model, a few price changes in key goods have had a heavy influence on overall inflation expectations, much like the wage and price controls in the United States in the 1970s, the administered prices in Brazil in the 2010s, and the food prices and electricity prices in South Africa in 2012. If so, then those past experiences suggest that the rise in expectations will be transitory and brief. One lesson from the 1967–1973 experience is that policymakers should keep a close eye on measures of the expected inflation anchor and not give in to the temptation to dismiss them as temporary noise or as vague psychological factors.

V. Conclusion

Every respectable central banker talks endlessly about credibility, steering expectations, and how well anchored inflation is. Yet, the experience of the Great Inflation in the United States illustrates that expectations are sometimes treated as mystical objects that policymakers infer from their wisdom and personal experiences. Economists also often refer to inflation expectations as an add-on factor to complement their model of fundamentals for why inflation rose or fell. Theorists assume that expectations are strictly rational and so need not be measured, while empiricists despair at how first-order moments from surveys have persistent biases and move so sluggishly.

This paper argued that we can measure inflation expectations in multiple ways. Looking deeper at household surveys, we can examine their distributions and especially the disagreement in them as measured by second- and third-order moments. Combining surveys of professionals with market prices provides valuable information to separate noise from signal. The wisdom of the crowd gives hints of what lies ahead. A parsimonious model can be used as a measurement tool to combine these hints into estimates.

During the Great Inflation, the expectations data show an anchor already drifting between 1967 and 1970, well before the end of Bretton Woods or the oil price shocks. This paper adds a new element to the bad measurement class of explanations of the Great Inflation: the bad measurement of inflation expectations. Not only did policymakers have incorrect estimates of productivity growth, the natural rate of unemployment, or the natural rate of interest, but they also failed to accurately measure expectations of inflation. These mistakes reinforced each other in leading them to miss the drift of the expected inflation and hence not preventing its loss. To the bad theories explanations, I add the observation that policymakers had no theory of expectations, appealing instead to vague animal spirits that led them to overly rely on transitory explanations for persistently rising inflation and to believing that wage and price controls sufficed. Finally, added to the bad luck stories is a new channel to the role of oil price shocks. The rise in the salient price of gas made usually sluggish inflation expectations jump quickly, and by a lot.

Confirming the usefulness of using inflation expectations data, the paper showed that they also detected the loss of the inflation anchor in Brazil in 2011–2016; that even if Turkey has not had a very firm anchor, and events are still too recent to be sure, the expectations data point to a significant change in 2018; that the expectations data do not give a false positive in South Africa 2010–2016, instead showing a steady expectations anchor in spite of several adverse shocks to actual inflation; and that expectations data also move consistently when an anchor is dropped as in the United States in 1979–1984. Finally, looking at current events, the data suggest that the anchor may have drifted in 2021, but it is still early enough that good policy or good luck in the near future can keep it in place.

Looking forward, there is much work to do in order to provide better measurements of the expected inflation anchor. The events of the 1960s provide a warning that even if a central bank cares about inflation expectations, trying to steer them using communication without a clear model or measurement of these expectations is a fruitless task. Ultimately, inflation

expectations are not an exogenous force but a mirror of the monetary and fiscal policy regimes being followed. Those policies are the fundamentals that anchor the expectations anchor itself.

ACKNOWLEDGMENTS I am grateful to Alan Blinder, Janice Eberly, Yuriy Gorodnichenko, and Jens Hilscher for comments. Discussions with Joao Ayres, Carlos Viana de Carvalho, Steve Davis, Erik Hurst, Ed Nelson, Robert Hetzel, and Valerie Ramey stimulated this project at its start. Gustavo Araujo, Viviane Bittencourt, Aloisio Campelo Junior, Pierre Siklos, and Justin Wolfers helped to collect some of the data series. I am grateful to Aleksandra Furmanek, Patrick Leitoff, Samantha Ong, and Annie Wang for their research assistance, and especially to Adrien Couturier, Rui Sousa, and Borui Niklas Zhu.

References

Alesina, Alberto. 1988. "Macroeconomics and Politics." *NBER Macroeconomics Annual* 3:13–62.

Alves, Sergio, Waldyr Areosa, and Carlos Carvalho. 2020. "Beating the 'Pros' with a Semi Structural Model of Their Own Inflation Forecasts." Paper presented at the 2020 World Congress of the Econometric Society, online, Bocconi University, Milan, Italy, and the 2018 NBER-NSF Time Series Conference, University of California, San Diego, September 7–8, 2018.

Andrade, Philippe, Olivier Coibion, Erwan Gautier, and Yuriy Gorodnichenko. 2020. "No Firm Is an Island? How Industry Conditions Shape Firms' Expectations." Working Paper 27317. Cambridge, Mass.: National Bureau of Economic Research.

Ang, Andrew, Geert Bekaert, and Min Wei. 2007. "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?" *Journal of Monetary Economics* 54, no. 4: 1163–212.

Angeletos, George-Marios, Zhen Huo, and Karthik Sastry. 2020. "Imperfect Macroeconomic Expectations: Evidence and Theory." *NBER Macroeconomics Annual* 35:1–86.

Ayres, Joao, Marcio Garcia, Diogo A. Guillén, and Patrick J. Kehoe. 2019. "The Monetary and Fiscal History of Brazil, 1960–2016." Working Paper 25421. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25421.

Binder, Carola Conces. 2016. "Estimation of Historical Inflation Expectations." *Explorations in Economic History* 61:1–31.

Blanchard, Olivier J. 1984. "The Lucas Critique and the Volcker Deflation." Working Paper 1326. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w1326/w1326.pdf.

Blinder, Alan S. 1982. "The Anatomy of Double-Digit Inflation in the 1970s." In *Inflation: Causes and Effects*, edited by Robert E. Hall. Chicago: University of Chicago Press.

Blinder, Alan S., and William J. Newton. 1981. "The 1971–1974 Controls Program and the Price Level: An Econometric Post-mortem." *Journal of Monetary Economics* 8, no. 1: 1–23.

Blinder, Alan S., and Jeremy B. Rudd. 2013. "The Supply-Shock Explanation of the Great Stagflation Revisited." In *The Great Inflation: The Rebirth of Modern Central Banking*, edited by Michael D. Bordo and Athanasios Orphanides. Chicago: University of Chicago Press.

Bonomo, Marco. 2018. "The Brazilian Economy in 2018: Diagnosis of a Crisis." Lecture given at the Center for Global Studies, University of Illinois Urbana-Champaign, September 11.

Bordo, Michael, and Barry Eichengreen. 2013. "Bretton Woods and the Great Inflation." In *The Great Inflation: The Rebirth of Modern Central Banking*, edited by Michael D. Bordo and Athanasios Orphanides. Chicago: University of Chicago Press.

Bureau for Economic Research. 2021. "The BER's Inflation Expectations Survey: Know Your Data." BER Research Note 3.

Burns, Arthur F. 1957. *Prosperity without Inflation*. New York: Fordham University Press.

Burns, Arthur F. 1987. "The Anguish of Central Banking." In *Money and the Economy: Central Bankers' Views*, edited by Pierluigi Ciocca. New York: Palgrave Macmillan Books.

Campbell, John Y., Robert J. Shiller, and Luis M. Viceira. 2009. "Understanding Inflation-Indexed Bond Markets." *Brookings Papers on Economic Activity*, Spring, 79–138.

Campbell, John Y., Adi Sunderam, and Luis M. Viceira. 2017. "Inflation Bets or Deflation Hedges? The Changing Risks of Nominal Bonds." *Critical Finance Review* 6, no. 2: 263–301.

Candia, Bernardo, Olivier Coibion, and Yuriy Gorodnichenko. 2020. "Communication and the Beliefs of Economic Agents." In *Economic Policy Symposium Proceedings: Navigating the Decade Ahead: Implications for Monetary Policy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Candia, Bernardo, Olivier Coibion, and Yuriy Gorodnichenko. 2021. "The Inflation Expectations of U.S. Firms: Evidence from a New Survey." Working Paper 28836. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28836.

Carroll, Christopher D. 2003. "Macroeconomic Expectations of Households and Professional Forecasters." *Quarterly Journal of Economics* 118, no. 1: 269–98.

Carvalho, Carlos, Stefano Eusepi, Emanuel Moench, and Bruce Preston. Forthcoming. "Anchored Inflation Expectations." *American Economic Journal: Macroeconomics.*

Castillo-Martinez, Laura, and Ricardo Reis. 2019. "How Do Central Banks Control Inflation? A Guide for the Perplexed." Working Paper. https://personal.lse.ac.uk/reisr/papers/99-perplexed.pdf.

Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy* 120, no. 1: 116–59.

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

de Leeuw, Frank, and Michael J. McKelvey. 1981. "Price Expectations of Business Firms." *Brookings Papers on Economic Activity*, no. 1: 299–314.

DeLong, J. Bradford. 1997. "America's Peacetime Inflation: The 1970s." In *Reducing Inflation: Motivation and Strategy*, edited by Christina D. Romer and David H. Romer. Chicago: University of Chicago Press.

Federal Reserve Board of Governors. 1968. Record of Policy Actions of the Federal Open Market Committee. Meeting minutes, November 26. https://www.federalreserve.gov/monetarypolicy/files/fomcropa19681217.pdf.

Federal Reserve Board of Governors. 1969. *56th Annual Report*. Washington. https://fraser.stlouisfed.org/title/annual-report-board-governors-federal-reserve-system-117/1969-2424.

Fischer, Stanley, and John Huizinga. 1982. "Inflation, Unemployment and Public Opinion Polls." *Journal of Money, Credit and Banking* 14, no. 1: 1–19.

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

Friedman, Milton, and Anna J. Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton, N.J.: Princeton University Press.

Gaglianone, Wagner Piazza, João Victor Issler, and Silvia Maria Matos. 2017. "Applying a Microfounded-Forecasting Approach to Predict Brazilian Inflation." *Empirical Economics* 53, no. 1: 137–63.

Gordon, Robert J. 1980. "Postwar Macroeconomics: The Evolution of Events and Ideas." *The American Economy in Transition*, edited by Martin Feldstein. Chicago: University of Chicago Press.

Grossman, Sanford J., and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *American Economic Review* 70, no. 3: 393–408.

Hetzel, Robert L. 1998. "Arthur Burns and Inflation." *Economic Quarterly* 84, no. 1: 21–44.

Hetzel, Robert L. 2020. "The Evolution of US Monetary Policy." In *Handbook of the History of Money and Currency*, edited by K. Yago, S. Battilossi, and Y. Cassis. Singapore: Springer.

International Monetary Fund. 2018. "Turkey: 2018 Article IV Consultation-Press Release; Staff Report; and Statement by the Executive Director for Turkey." Country Report 18/110. Washington: International Monetary Fund.

Jalil, Andrew J., and Gisela Rua. 2016. "Inflation Expectations and Recovery in Spring 1933." *Explorations in Economic History* 62:26–50.

Juster, F. Thomas, and Robert Comment. 1980. "A Note on the Measurement of Price Expectations." Working Paper.

Juster, F. Thomas, Paul Wachtel, Saul Hymans, and James Duesenberry. 1972. "Inflation and the Consumer." *Brookings Papers on Economic Activity*, no. 1: 71–121.

Kabundi, Alain, Eric Schaling, and Modeste Some. 2015. "Monetary Policy and Heterogeneous Inflation Expectations in South Africa." *Economic Modelling* 45:109–17.

Kara, Hakan. 2021. "Turkey's Ceaseless Battle with Price Stability." Lecture, LSE SU Developing Markets Society, March 15, available on YouTube, https://www.youtube.com/watch?v=GWJCCnl7_bU.

Knotek, Edward S., Raphael S. Schoenle, Alexander M. Dietrich, Keith Kuester, Gernot J. Müller, Kristian Ove R. Myrseth, and Michael Weber. 2020. "Consumers and Covid-19: A Real-Time Survey." Federal Reserve Bank of Cleveland Economic Commentary 2020-08. https://www.clevelandfed.org/en/newsroom-and-events/publications/economic-commentary/2020-economic-commentaries/ec-202008-consumers-and-covid19.aspx.

Mankiw, N. Gregory, and Ricardo Reis. 2010. "Imperfect Information and Aggregate Supply." In *Handbook of Monetary Economics, Vol. 3A*, edited by Benjamin Friedman and Michael Woodford. Amsterdam: North Holland.

Mankiw, N. Gregory, and Ricardo Reis. 2018. "Friedman's Presidential Address in the Evolution of Macroeconomic Thought." *Journal of Economic Perspectives* 32, no. 1: 81–96.

Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers. 2004. "Disagreement about Inflation Expectations." *NBER Macroeconomics Annual* 18:209–48.

Martin, William McChesney. 1959. "A Year of Recession and Recovery." *Federal Reserve Bulletin* 45, no. 2 (February): 110–19.

McCulloch, J. Huston. 1980. "The Ban on Indexed Bonds, 1933–77." *American Economic Review* 70, no. 5: 1018–21.

Meltzer, Allan H. 2005. "Origins of the Great Inflation." *Federal Reserve Bank of St. Louis Review* 87 (March): 145–176.

Meltzer, Allan H. 2010. *A History of the Federal Reserve, Volume 2, Book 1, 1951–1969*. Chicago: University of Chicago Press.

Mminele, Daniel. 2019. "Understanding the Core Drivers of Inflation in South Africa." Keynote address at S&P Dow Jones Indices South African Seminar, February 12. https://www.bis.org/review/r190212e.pdf.

Nelson, Edward. 2005. "The Great Inflation of the Seventies: What Really Happened?" *B.E. Journal of Macroeconomics* 5, no. 1: 1–50.

Nelson, Edward. 2020. *Milton Friedman and Economic Debate in the United States, 1932–1972.* Chicago: University of Chicago Press.

Okun, Arthur M. 1970. *The Political Economy of Prosperity*. New York: W. W. Norton.

Organisation for Economic Cooperation and Development (OECD). 2018. *OECD Economic Surveys: Turkey 2018*. Paris: OECD.

Orphanides, Athanasios. 2003. "The Quest for Prosperity without Inflation." *Journal of Monetary Economics* 50, no. 3: 633–63.

Pesaran, M. Hashem, and Martin Weale. 2006. "Survey Expectations." In *Handbook of Economic Forecasting, Vol. 1*, edited by G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: North Holland.

Pick, Franz. 1971. *Pick's Currency Yearbook*. New York: Pick Publishing Co.

Primiceri, Giorgio E. 2006. "Why Inflation Rose and Fell: Policy-Makers' Beliefs and U.S. Postwar Stabilization Policy." *Quarterly Journal of Economics* 121, no. 3: 867–901.

Reid, Monique, Stan Du Plessis, and Pierre Siklos. 2021. "What Drives Household Inflation Expectations in South Africa? Demographics and Anchoring under Inflation Targeting." *Economic Systems* 45, no. 3: article 100878.

Reis, Ricardo. 2003. "Where Is the Natural Rate? Rational Policy Mistakes and Persistent Deviations of Inflation from Target." *Advances in Macroeconomics* 3, no. 1: 1–38.

Reis, Ricardo. 2021. "The People versus the Markets: A Parsimonious Model of Inflation Expectations." Discussion paper 15624. London: Centre for Economic Policy Research.

Romer, Christina D. 2005. Commentary on "Origins of the Great Inflation." *Federal Reserve Bank of St. Louis Review*, March/April 2005, Part 2, 177–86.

Romer, Christina D., and David H. Romer. 1989. "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz." *NBER Macroeconomics Annual* 4:121–70.

Romer, Christina D., and David H. Romer. 2002. "The Evolution of Economic Understanding and Postwar Stabilization Policy." In *Economic Policy Symposium Proceedings: Rethinking Stabilization Policy*. Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City.

Sargent, Thomas J. 1999. *The Conquest of American Inflation*. Princeton, N.J.: Princeton University Press.

Tobin, James. 1980. "Stabilization Policy Ten Years After." *Brookings Papers on Economic Activity*, no. 1: 19–71.

US President and Council of Economic Advisers. 1969. *Economic Report of the President together with the Annual Report of the Council of Economic Advisers*. Washington: United States Government Printing Office. https://fraser.stlouisfed.org/title/economic-report-president-45/1969-8140.

Val, Flávio, and Gustavo Araujo. 2019. "Breakeven Inflation Rate Estimation: An Alternative Approach Considering Indexation Lag and Seasonality." Working Paper 493. Brasilia: Central Bank of Brazil.

Weise, Charles L. 2012. "Political Pressures on Monetary Policy during the US Great Inflation." *American Economic Journal: Macroeconomics* 4, no. 2: 33–64.

Comments and Discussion

COMMENT BY

ALAN S. BLINDER I have four main points to make about Ricardo Reis's fascinating paper. The first is a point of grammar: The word *anchor* can be a noun or a verb. While Ricardo concentrates on the noun form, I think the verb form is the more interesting and important one. Second, while actual inflation and expected inflation are jointly endogenous variables with each one influencing the other, just as Reis says, the direction in which causation is quantitatively more important matters. Is it mainly expected inflation that drives actual inflation, which comes close to Reis's view? Or is it mainly the actual experience with inflation that moves expected inflation, which comes closer to my own view? Third, readers should realize what a difficult task Reis has set for himself in this paper. I'm tempted to call it an impossible task, but Reis has a track record of making progress on nearly impossible problems—and he does so here. Nonetheless, the extreme difficulty of the task limits how convincing his conclusions can be. In particular, and finally, I take issue with Reis's "bottom line" story of how and when the United States lost its inflation anchor in the 1967–1970 period.

NOUNS AND VERBS When the word *anchor* is used as a noun, it refers to a physical object. That's the metaphor that Reis uses repeatedly: the ship of inflation is either anchored in the seabed or adrift. But when *anchor* is used as a verb, it's about how something—in this case, expected inflation—gets anchored.

Reis concentrates on the noun, specifically on estimating a numerical value for long-run expected inflation. But to me, and I think to macroeconomics in general, the verb "to anchor" is more important. It's about how the central bank is supposed to hold actual and expected inflation at or near whatever target it selects.

15985-05b-Reis_Comm&GD-4thPgs.indd 362 5/12/22 12:07 PM

Many methods of anchoring inflationary expectations have been proposed and used over the years, dating back at least to the gold standard. Some methods, like the gold standard, rely on a fixed (or almost fixed) exchange rate to anchor expectations. That includes, of course, the Bretton Woods system and various sorts of exchange rate pegs. In recent history, episodes in which the Argentine peso or the Brazilian real were tied to the US dollar come to mind.[1] Each of these pegs worked—until they didn't. The European Monetary Union was also thought of as a way to bring, say, Italian or Spanish inflation rates down to German levels.

Milton Friedman and the monetarists, of course, offered the k-percent rule for money growth as an attractive way to anchor the inflation rate. After all, inflation is always and everywhere a monetary phenomenon, right? Well, apparently not. Almost no one advocates a k-percent rule these days.

In more recent decades, attention has shifted toward inflation targeting, or its first cousin, price-level targeting. In either case, the stated inflation target is supposed to anchor inflationary expectations. Inflation targeting has certainly scored considerable successes, both in the academy and in the real world of central banking. It's a live candidate today.

One open, but not terribly important, question is whether following a Taylor rule for monetary policy is either an operational instruction for implementing a policy of inflation targeting or a different way of anchoring inflationary expectations to the target.

Finally, I'd like to add something I call "earned credibility" to the list. I'm thinking about anchoring episodes like the Bundesbank standard in pre-euro Germany, the Thatcher standard in the United Kingdom, and the Volcker standard in the United States. Market participants understood, or quickly learned, that the Bundesbank or Margaret Thatcher or Paul Volcker would not tolerate much inflation, and that belief anchored expectations. These episodes are, of course, first cousins of Kenneth Rogoff's (1985) conservative central banker.

One reading of Reis's major conclusion about the rise of inflation in the United States in the 1960s and 1970s is that the Bretton Woods system did not effectively discipline the United States and that we never had a Martin standard.

For a system to anchor inflationary expectations effectively, its design must satisfy certain requirements. First, it must be feasible to implement—the

1. See, for example, Spiegel (2002) on Argentina and Gruben and Kiser (1999) on Brazil.

central bank must actually be able to do it. That consideration probably doomed the k-percent rule decades ago.

Second, the mechanism must be effective in pinning down the expected rate of inflation, which was perhaps one major reason why the Federal Reserve abandoned conventional inflation targeting in 2020 in favor of what it calls *flexible average inflation targeting*. The Federal Reserve clearly was not achieving its 2 percent target and became worried that expected inflation was drifting downward.

Third, the system for anchoring expectations should be robust to inflationary shocks, which is certainly one reason to focus on core inflation rather than headline inflation. This problem is plaguing the Federal Reserve and other central banks right now as bottlenecks have driven the inflation rate to heights not seen in decades.

Finally, and perhaps most important, the system chosen to anchor inflationary expectations must be credible: markets, businesses, and households must believe it will work. Credibility in this context comes in two main parts: the central bank must first achieve credibility through its words and then retain credibility by its actions. Volcker did both for the Federal Reserve. Arthur Burns did neither.

These four criteria interact in obvious ways, of course. For example, expectations won't be credible if the method of anchoring is infeasible or doesn't work.

EXPECTATIONAL DOMINANCE As noted already, and as mentioned by Reis, there is two-way causation between actual and expected inflation. But which direction is quantitatively dominant?

Much of modern macroeconomics sees expected inflation as the dominant factor, with actual inflation trundling along passively behind it. For example, the core of many New Keynesian models looks something like this:

$$y_t = E_t y_{t+1} - \sigma\left(i_t - E_t \pi_{t+1} - r_t^n\right) \quad \text{(AD)}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa y_t \quad \text{(AS)}$$

Aggregate demand depends on the gap between the natural and actual real interest rates, the latter of which depends on the expected inflation rate. What passes for the aggregate supply curve sees expected inflation driving actual inflation rather than the other way around. Notice that lagged inflation is nowhere to be found. And needless to say, the expectations operator in both equations connotes rational expectations.

15985-05b-Reis_Comm&GD-4thPgs.indd 364 5/12/22 12:07 PM

A well-known and extremely clever paper by Jonathon Hazell, Juan Herreño, Emi Nakamura, and Jón Steinsson (2020) uses a model like this to derive the following solution for this period's inflation rate:

$$\pi_t = \phi \tilde{u}_t + E_t \pi_{t+\infty} + \omega_t,$$

where $\tilde{u}_t$ is a complicated unemployment variable. The key point to notice here is that the inflation rate expected "at infinity" drives the current inflation rate. Hazell and others (2020) use this equation to explain why it is so hard to move current inflation: long-run inflationary expectations are too well anchored. Notice that this equation implies that a 1 percent increase in π at infinity will move current inflation by 1 percent immediately. Really?

I'll risk being banished from the tribe by suggesting that perhaps this version of the inflationary process, which has become dominant in academia, goes overboard in its view of the (overwhelming) importance of long-run expectations. Objections to the pure rational expectations model are well known and numerous: nominal contracts, inertia, inattention (whether rational or not), and more. These objections are typically ignored in many New Keynesian models. And without rational expectations, you don't get the inflation equation just above.

There is, of course, an alternative hypothesis to rational expectations: that expected inflation normally reacts to, and therefore naturally lags behind, actual inflation. The basic idea here is that the expected inflation rate is less a leader and more a follower: people come to expect higher inflation if and when they actually see higher inflation. If that's the case, the inflation process is likely to be more inertial than in typical New Keynesian models, and pronouncements about future inflation have less leverage over inflationary expectations than the actual behavior of inflation in real time. Adaptive expectations are one well-known example of such an inertial process, but there is nothing sacred about that particular formulation.

Reis's paper concentrates on a different departure from rational expectations: that different types of agents have different expectations. In the pure rational expectations world, this is impossible because all agents have the same information, know how to process it efficiently, and have the same model in their heads (or on their computers). In the real world, by contrast, different expectations are not just possible, they are an obvious reality. Reis cleverly extracts useful information from the differences between the expectations of professionals and those of ordinary people (in the Michigan survey). The complex details of how he does this are in his earlier paper

(Reis 2021), and I won't discuss them here. Suffice it to say that he pulls a pretty interesting rabbit out of a pretty deep hat.

But then Reis strains to squeeze virtually everything into the Procrustean bed of inflationary expectations. (He is, after all, the very model of a modern macroeconomist.) Supply shocks, political manipulation of monetary policy by Nixon and Burns, and the Federal Reserve's judgmental errors are all pushed into this bed. Strange bedfellows, I'd say. Arguing that inflation is always and everywhere an expectational phenomenon is no better than arguing that inflation is always and everywhere a monetary phenomenon. Yes, expectations matter. But so do other things.

A QUIXOTIC GESTURE? Readers should realize that Reis set himself an almost impossible task in this paper, making it unsurprising that his conclusions are not entirely convincing. The wonder is that he made any progress at all, given that (in his words) "between 1970 and 1995, many anchors were lost, but there are almost no expectations data; between 1995 and 2020, there are data, but no lost anchors." For the period on which he concentrates, the late 1960s and early 1970s in the United States, he is forced to cobble together what expectational data he can, even though the time horizon in the survey data "was for the most part one year" while "economic theory would suggest that it is longer-horizon inflation expectations that provide the anchor."

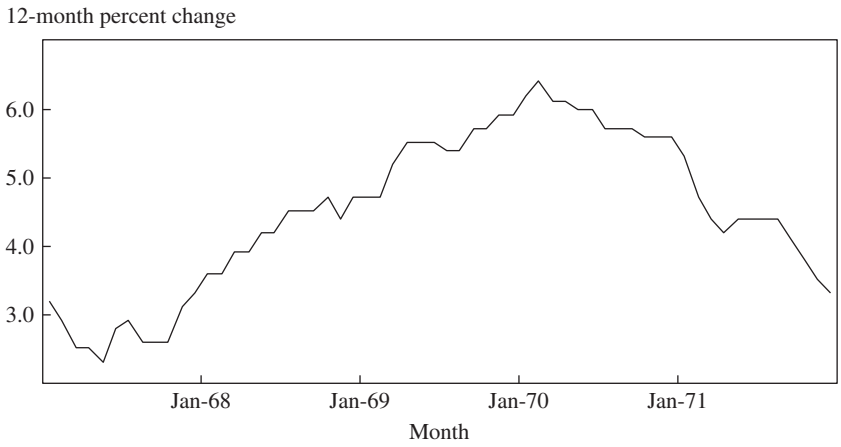
A less intrepid researcher might have given up right there. But instead, Reis makes creative, even ingenious, efforts to minimize these problems. I certainly don't claim that I could do better. But in the end, is the story believable?

THE BEGINNINGS OF THE GREAT INFLATION Reis's figure 10, panel A, summarizes his main conclusion about when the United States lost its inflation anchor. As he emphasizes, "the expectations data show an anchor already drifting between 1967 and 1970, well before the end of Bretton Woods or the oil price shocks." According to his estimate, the expectational anchor drifted up by 2.5 percentage points between 1967 and 1970. (But then it went down in 1971.) Is that right? Did the United States really lose its anchor that early, rather than in 1972–1974, as many scholars would say?

Figure 1 shows the CPI inflation rate (measured from twelve months previously) over the same time span as Reis's figure 10, panel A. Allowing for the imperfect alignment in time—Reis's numbers are annual while the CPI data are monthly—there is a striking resemblance between the two figures. In a word, Reis's constructed anchor echoes actual inflation. So let's review the conventional story of the lead-up to the Great Inflation.

15985-05b-Reis_Comm&GD-4thPgs.indd 366 5/12/22 12:07 PM

Figure 1. CPI Inflation (Twelve-Month Percentage Change), 1967–1971



Source: Bureau of Labor Statistics.

Remember, I'm testing the parsimonious model expected inflation = *f* (recent actual inflation).

In 1965, the US economy appeared to be just about at full employment with very low inflation. Then President Lyndon Johnson ordered the first stages of the military buildup for the Vietnam War. His economic advisers, led first by Walter Heller and then by Arthur Okun, warned him that layering more aggregate demand atop a fully employed economy would cause inflation. And they urged him to raise taxes to pay for the war, to avoid overheating. Johnson refused, however; he wanted both guns *and* butter. Perhaps even more important, LBJ did not want tax hikes to make the costs of the war more visible to the American public (Okun 1970).

But the Federal Reserve perceived the inflationary danger, and it raised interest rates for about a year, from late 1965 until late 1966. These rate hikes enraged Johnson, who famously summoned William McChesney Martin down to Texas to barbecue him (only figuratively!). Martin produced a notable (and contemporaneously noted) credit crunch; and precipitated what was then called a "growth recession," a slowdown in economic growth that helped bring down inflation.[2] Yet Reis is critical of the Federal Reserve: "the monetary restraint was gone by the end of 1966."

2. The Martin-Johnson story is nicely told by Kevin Granville (2017).

True, but I wouldn't be quite so dismissive. Martin was, after all, defying a strong-willed president at a time when the Federal Reserve considered itself part of an economic team led by the administration. Furthermore, and amazingly to the modern eye, control of inflation was then considered more the responsibility of fiscal policy than of monetary policy. Here's a remarkable quotation from the 1968 *Economic Report of the President*: "It has been and remains the conviction of *both the Administration and the Federal Reserve System* that the Nation should depend on *fiscal policy, not monetary policy*, to carry the main burden of the additional restraint on the growth of demand that now appears necessary for 1968" (Council of Economic Advisers 1968; emphasis added). A statement like that, explicitly endorsed by the Federal Reserve, is simply unimaginable today.

So did fiscal policy carry the burden? Well, Johnson was finally convinced of the dangers of rising inflation by January 1967, at which time he recommended an income tax surcharge. Congress, however, required a lot more convincing; it didn't enact the surtax until June 1968. By then, inflation was above 4 percent and rising (Okun 1970).

The Federal Reserve didn't wait for Congress to act. Beginning in the fall of 1967, it embarked on a tightening cycle that would eventually raise the funds rate about 500 basis points in under two years. As the CPI graph (figure 1) makes clear, the inflation rate began to decline early in 1970. Reis condemns the Federal Reserve for being insufficiently aggressive in the 1966–1968 period—when the anchor started slipping away. Maybe so. But it seems to me that tight money played a major role in bringing inflation down from about 6 percent to about 3.5 percent *before* President Richard Nixon instituted wage-price controls in August 1971.

Was the anchor (as a noun) lost by 1970? I'm not as convinced as Reis is. But, more important, there was no anchoring (a verb) at all. The United States would not acquire any until we went on the Volcker standard in the early 1980s.

That said, inflation did soar as price controls were being lifted in 1973–1974 (Blinder and Newton 1981). The large food price shocks of 1972 and the stunning oil price shock of 1973–1974 were other major reasons, as Reis correctly notes. He also correctly observes that inflation did *not* revert to 1970–1971 levels once these large shocks passed through the system. With nothing to anchor inflation, inertia took over—until Volcker arrived at the Federal Reserve.

The Volcker standard was inherited by Alan Greenspan (becoming the Greenspan standard) and then by Ben Bernanke, who preferred to depersonalize and institutionalize the anchor by adopting inflation targeting.

15985-05b-Reis_Comm&GD-4thPgs.indd 368 5/12/22 12:07 PM

The Federal Reserve's more recent adoption of flexible average inflation targeting is the latest variation on that theme. But it is far too early to judge its success.

Please notice, however, that this entire conventional story about rising inflation makes no mention of expected inflation driving actual inflation. If I may apply Laplace's famous statement, made regarding a far weightier matter, I have no need of that hypothesis.

REFERENCES FOR THE BLINDER COMMENT

Blinder, Alan S., and William J. Newton. 1981. "The 1971–1974 Controls Program and the Price Level: An Econometric Post-mortem." *Journal of Monetary Economics* 8, no. 1: 1–23.

Council of Economic Advisers. 1968. *Economic Report of the President: Hearings before the Joint Economic Committee, Congress of the United States.* Washington: Government Printing Office.

Granville, Kevin. 2017. "A President at War with His Fed Chief, 5 Decades before Trump." *New York Times*, June 13.

Gruben, William C., and Sherry Kiser. 1999. "Brazil: The First Financial Crisis of 1999." *Southwest Economy*, Federal Reserve Bank of Dallas, March/April. https://www.dallasfed.org/~/media/documents/research/swe/1999/swe9902c.pdf.

Hazell, Jonathon, Juan Herreño, Emi Nakamura, and Jón Steinsson. 2020. "The Slope of the Phillips Curve: Evidence from U.S. States." Working Paper 28005. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28005.

Okun, Arthur M. 1970. *The Political Economy of Prosperity*. New York: W. W. Norton.

Reis, Ricardo. 2021. "The People versus the Markets: A Parsimonious Model of Inflation Expectations." Discussion Paper 15624. London: Centre for Economic Policy Research.

Rogoff, Kenneth S. 1985. "The Optimal Degree of Commitment to an Intermediate Monetary Target." *Quarterly Journal of Economics* 100, no. 4: 1169–89.

Spiegel, Mark. 2002. "Argentina's Currency Crisis: Lessons for Asia." *Economic Letter*, Federal Reserve Bank of San Francisco, August 23. https://www.frbsf.org/economic-research/publications/economic-letter/2002/august/argentina-currency-crisis-lessons-for-asia/.

COMMENT BY

YURIY GORODNICHENKO The Great Inflation of the 1970s left many enduring marks on macroeconomic thinking and policy. For example, inflation expectations moved from relative obscurity to a key element for policymaking. But what determines inflation expectations? How should we

measure inflation expectations? Whose expectations are important? What is the effect of inflation expectations on the economy? How can central bankers use inflation expectations for policy? Although much progress has been made to shed light on these issues, many questions remain open. Historically, there has been strong demand from the central banking community to better understand the interplay between inflation and inflation expectations.[1] But there is a new sense of urgency to get answers given the current developments in the United States and other advanced economies. Indeed, with elevated inflation and inflation expectations, some observers and commentators are concerned that we are on a path to repeat the experience of the 1970s (Irwin 2021). Reis's analysis is thus most timely and highly policy relevant.

His work makes several important contributions. First, after unearthing historical data on inflation expectations for various economic players, Reis documents that inflation expectations started to become unanchored circa 1967, which is well before the time suggested by other analyses. His timing suggests an important role of information rigidities and the credibility of the central bank. Intuitively, if we start in a low-inflation environment and a credible central bank, the public pays little attention to inflation and inflation expectations are relatively insensitive to inflationary shocks. As a result, it takes time for the public to accumulate enough observations to become concerned about inflation and raise their inflation expectations. This also means that the central bank can "spend" its credibility/inattention capital on addressing problems in the economy without igniting inflation concerns but the credibility/inattention capital gives only a temporary space for policy maneuvers. This dynamic contrasts with the credible disinflations where economic players pay attention to inflation and thus revise their (unanchored) inflation expectation quickly in response to incoming data and policy announcements. Second, building on Mankiw, Reis, and Wolfers

1. The following quotes should provide a sense of what the central banking community thinks about inflation expectations. Greenspan (1994) asserted, "I am not saying what [inflation expectations are] a function of. We know it's a very difficult issue, but that is the key variable. It's important, but just because we can't make a judgment as to what these driving forces are in an econometric sense doesn't mean that it's not real." Bernanke (2007) observed, "How should we measure inflation expectations, and how should we use that information for forecasting and controlling inflation? I certainly do not have complete answers to those questions, but I believe that they are of great practical importance. . . . Information on the price expectations of businesses—who are, after all, the price setters in the first instance . . . is particularly scarce." Yellen (2016) noted, "Perhaps most importantly, we need to know more about the manner in which inflation expectations are formed and how monetary policy influences them."

15985-05b-Reis_Comm&GD-4thPgs.indd 370 5/12/22 12:07 PM

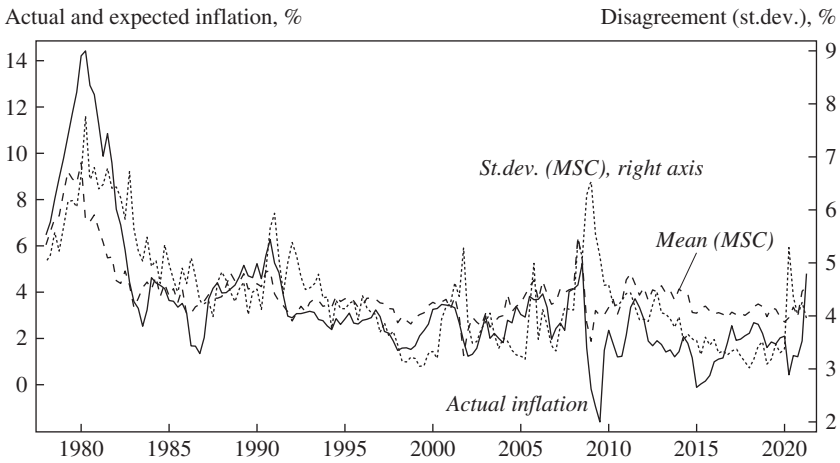
(2004), Reis proposes a useful, real-time indicator for how (un)anchored inflation expectations are. Specifically, he shows that when the right tail of the cross-sectional distribution of inflation expectations starts to increase, one may have an early warning that inflation expectations could be getting unanchored. Because disagreement is largely driven by the right tail of the distribution, one can also use disagreement as an early warning indicator. Third, Reis provides a new perspective on why the Great Inflation happened. In particular, he argues that inflation expectations were poorly understood and measured. For example, inflation expectations were reduced to ad factoring (i.e., unexplained wedges) in macroeconomic models. In a similar spirit, policymakers talked about "inflation psychology" rather than relying on proper measurement of inflation expectations. These factors exacerbated other problems such as poor measurement of output gap (Orphanides 2001), perceived inability of the Federal Reserve to control inflation (Romer and Romer 2013), and energy price hikes (Hamilton 1983) that led to high inflation. Finally, Reis draws some worrying parallels between the 1970s and the current situation.

I find Reis's insightful, detailed analysis convincing and helpful for thinking about the rise of the Great Inflation as well as current inflation developments. At the same time, I have a more positive outlook for future inflation, although obviously there is huge uncertainty in any forecast given COVID-19 vagaries. There are several reasons for why we are unlikely to have a repeat of the 1970s and currently high inflation will likely turn out to be transitory.[2]

First, it is true that households' inflation expectations are high now and there is much disagreement about future inflation. This is a source of concern because the same constellation was characteristic of the Great Inflation. However, it is not unusual to have both high mean and high disagreement (figure 1). For example, households' inflation expectations and disagreement increased significantly during the inflation scare of the early 1990s. Yet, an increase in inflation during this episode turned out to be transitory. In a similar spirit, inflation expectations and disagreement ran high in the years preceding the Great Recession but inflation did not become a chronic problem. We also observe that the inflation scare of 2011–2012 had elevated inflation expectations and disagreement but no systematically high inflation emerged. These episodes suggest that rising disagreement can be

2. To keep this discussion related to Reis's work, I will focus on inflation expectations, but there are obviously many other factors to keep in mind. For example, cost of living adjustment (COLA) clauses are not as prevalent in labor contracts now as they used to be during the Great Inflation.

Figure 1. Inflation Expectations in the Michigan Surveys of Consumers



Source: Michigan Surveys of Consumers.

Note: The figure plots time series of actual inflation (CPI, year-on-year), one-year-ahead mean expected inflation in the Michigan Surveys of Consumers, and disagreement (standard deviation) for expected inflation in the survey. Author's calculations.

a useful leading indicator of unanchored inflation expectations and high future inflation but like any other leading indicator it can generate false alarms. Interestingly, in each of these episodes there was much talk about runaway inflation and debasing the dollar and so the inflation scares like the one we have today are rather familiar (Chan 2011).

Second, the Federal Reserve raised interest rates in the 1990s and 2000s thus possibly averting problems with inflation—and thus the alarms could appear false because of the policy response—but the Federal Reserve did not raise interest rates in 2011–2012 and this later episode is likely more informative for understanding the current environment than the 1970s. Indeed, similar to 2011–2012, the Federal Reserve has interest rates at the zero lower bound, the economy is recovering after a major crisis, there is much underemployment, and energy prices are high. On the other hand, fiscal policy appears to be more expansionary now but, at the time of the writing, it remains to be seen whether fiscal support will be withdrawn quickly as was done after the Great Recession. Importantly, inflation expectations of households stayed high well after 2011—more on this shortly—but the hike in actual inflation was short-lived. In fact, the economy struggled afterward with persistently low inflation, which is consistent with disinflationary pressures due to massive, persistent underemployment after the Great

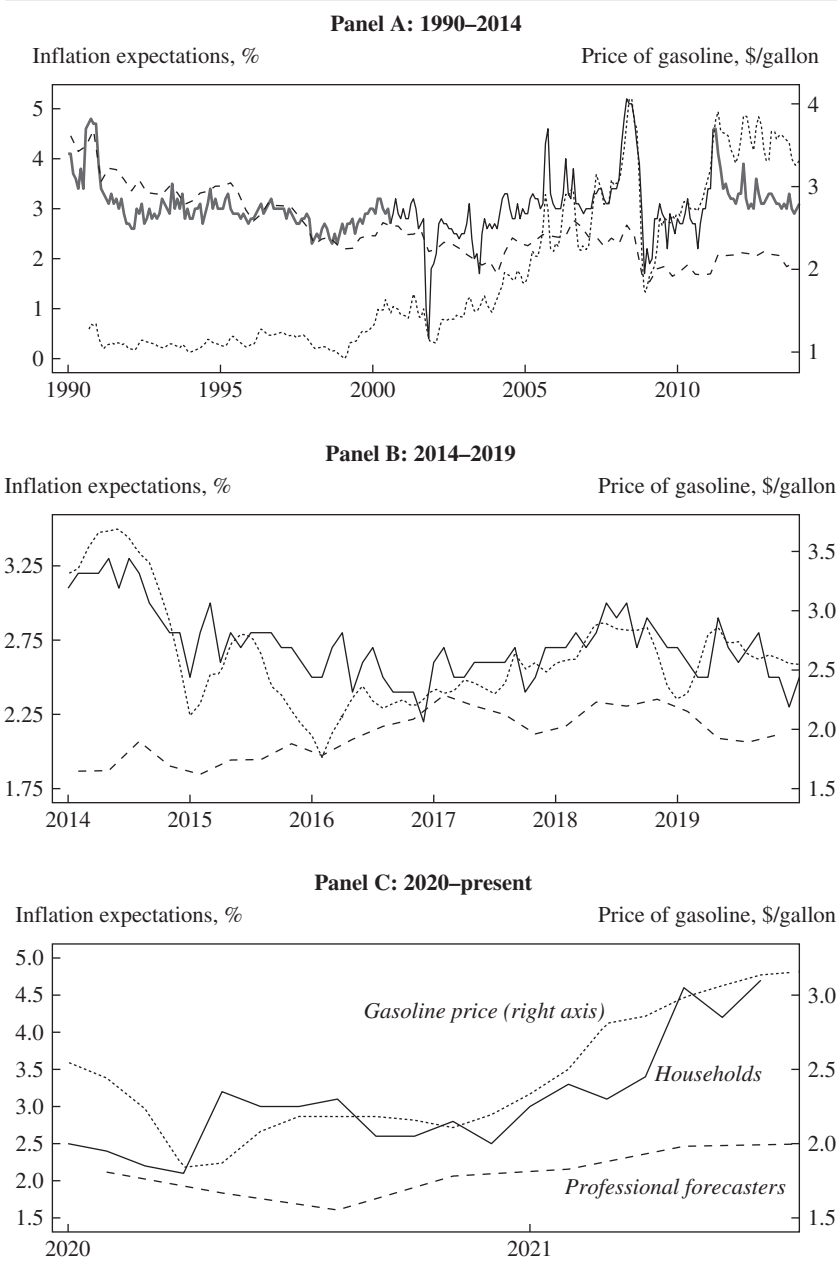
15985-05b-Reis_Comm&GD-4thPgs.indd 372 5/12/22 12:07 PM

Recession via the Phillips curve. In other words, although various shocks could have raised inflation and inflation expectations in 2011, the systematic disinflation force dominated the longer-run dynamics. Given that current employment is well below the prepandemic level (at present, the employment to population ratio is roughly at the level that was observed at the trough of the Great Recession), one may project that the same systematic force will weigh down on inflation in the coming years.

Third, inflation expectations of households are remarkably sensitive to changes in energy prices. Panel A of figure 2 shows that, since the early 1990s, households' inflation expectations track the price of gasoline closely. Note that in this relationship it is the *level* of gasoline prices rather than the *change* in gasoline price that matters for what people think about future inflation. Panel B of figure 2 illustrates that, although potentially evolving over time, this relationship applies to recent prepandemic years as well when policymakers were concerned about a possibly overheating economy. When oil prices collapsed in 2014, households revised their inflation expectations downward. Panel C focuses on the COVID-19 crisis and documents that again households' inflation expectations and the price of gasoline co-move strongly. On the other hand, professional forecasters have inflation expectations with weak sensitivity to energy prices, and they see little chance of high inflation on the horizon. Why would households—and likely firms, although there is more uncertainty here given the dearth of high-quality surveys of business executives and managers—be so reactive to the price of gasoline? One may interpret this empirical pattern as a sign of success: by delivering low, stable inflation for many years, the Federal Reserve made inflation an uninteresting subject to the general public. Consistent with this view, surveys find that the public is largely unaware of monetary policy (Binder 2017; Lamla and Vinogradov 2019; Coibion, Gorodnichenko, and Weber 2019). Instead, the public appears to use salient prices of frequently purchased, relatively homogenous goods as a shortcut for forming their inflation expectations (Coibion and Gorodnichenko 2015; Cavallo, Cruces, and Perez-Truglia 2017; D'Acunto and others 2021). In this case, inflation expectations of households could not only depart materially from the predictions of rational and well-informed agents like professional forecasters but also become more volatile and sensitive to short-term shocks that drive energy prices. For example, in 2008, energy prices shot up and household inflation expectations increased by 2 percentage points. This ignited familiar talk about the return of the 1970s. But this increase in expectations reversed itself as soon as gasoline prices fell a few months later. Hence, to the extent energy markets experience transitory difficulties

15985-05b-Reis_Comm&GD-4thPgs.indd 373 5/12/22 12:07 PM

Figure 2. Inflation Expectations and the Price of Gasoline



Sources: Michigan Surveys of Consumers, https://data.sca.isr.umich.edu/tables.php; Survey of Professional Forecasters, https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/survey-of-professional-forecasters; and Federal Reserve Economic Data, https://fred.stlouisfed.org/.

15985-05b-Reis_Comm&GD-4thPgs.indd 374 5/12/22 12:07 PM

now, one may also predict that households' inflation expectations will abate in the future.

Fourth, the mapping from inflation expectations to actions is likely to be more nuanced than posited by mainstream models. Specifically, the standard New Keynesian framework predicts that increased inflation expectations should stimulate current consumption as households substitute intertemporally. But if current consumption increases and hence raises the cost of producing goods, inflation expectations should rise further, which in turn spurs another round of increased consumption. This spiral is particularly dangerous for an already overheated economy, thus prompting the central bank to step on the brakes at the first signs of rising future inflation. According to this account, inflation should be associated with economic booms but households have a stagflationary view of the world (Kamdar 2018): they associate high inflation with high unemployment. Consistent with this stagflationary view, randomized controlled trials (Coibion and others 2019; Coibion, Gorodnichenko, and Weber 2019) find that exogenously raised inflation expectations of households lead to less frequent purchases of durable goods. As a result, while households act on their inflation expectations, the inflation-spending spiral appears to be a weaker propagation force than thought before. Furthermore, as the economy improves, one may predict that households will revise their inflation expectations down, thus further alleviating concerns about runaway inflation.

Does this mean we can't relive the 1970s? In the famous words of Yogi Berra, "it's tough to make predictions, especially about the future," which is particularly relevant in light of COVID-19 uncertainties. But we can learn from the past mistakes and the 1970s taught us a number of lessons. Few central bankers now believe that inflation is outside their control. Measurement of inflation expectations has improved dramatically. Macroeconomic theory made great strides in incorporating and modeling inflation expectations. Of course, we do not have complete answers but we know enough to not step on the Great Inflation rake again. In my view, the main risk now is a premature withdrawal of fiscal/monetary support for the recovering economy, a mistake that inflicted unnecessary pain in the aftermath of the Great Recession (Coibion, Gorodnichenko, and Koustas 2013).

REFERENCES FOR THE GORODNICHENKO COMMENT

Bernanke, Ben S. 2007. "Inflation Expectations and Inflation Forecasting." Speech at the Monetary Economics Workshop of the National Bureau of Economic Research Summer Institute, Cambridge, Massachusetts, July 10. https://www.federalreserve.gov/newsevents/speech/bernanke20070710a.htm.

Binder, Carola Conces. 2017. "Fed Speak on Main Street: Central Bank Communication and Household Expectations." *Journal of Macroeconomics* 52, no. C: 238–51.

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

Chan, Sewell. 2011. "In Congress, Bernanke Faces Questions about Inflation." *New York Times*, February 9. https://www.nytimes.com/2011/02/10/business/economy/10fed.html.

Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, and Maarten van Rooij. 2019. "How Does Consumption Respond to News about Inflation? Field Evidence from a Randomized Control Trial." Working Paper 26106. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w26106.

Coibion, Olivier, and Yuriy Gorodnichenko. 2015. "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation." *American Economic Journal: Macroeconomics* 7, no. 1: 197–232.

Coibion, Olivier, Yuriy Gorodnichenko, and Dmitri Koustas. 2013. "Amerisclerosis? The Puzzle of Rising U.S. Unemployment Persistence." *Brookings Papers on Economic Activity*, Fall, 193–260.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber. 2019. "Monetary Policy Communications and Their Effects on Household Inflation Expectations." Working Paper 25482. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25482.

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

Greenspan, Alan. 1994. Federal Open Market Committee Meeting Minutes, July 5–6.

Hamilton, James D. 1983. "Oil and the Macroeconomy since World War II." *Journal of Political Economy* 91, no. 2: 228–48.

Irwin, Neil. 2021. "How 10 Prominent Economists Think about Overheating." *New York Times*, March 24. https://www.nytimes.com/2021/03/24/upshot/economy-inflation.html.

Kamdar, Rupal. 2018. "The Inattentive Consumer: Sentiment and Expectations." Manuscript. https://rupalkamdar.github.io/pdfs/Inattentive_Consumer.pdf.

Lamla, Michael J., and Dmitri V. Vinogradov. 2019. "Central Bank Announcements: Big News for Little People?" *Journal of Monetary Economics* 108, no. C: 21–38.

Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers. 2004. "Disagreement about Inflation Expectations." *NBER Macroeconomics Annual* 18:209–70.

Orphanides, Athanasios. 2001. "Monetary Policy Rules Based on Real-Time Data." *American Economic Review* 91, no. 4: 964–85.

Romer, Christina D., and David H. Romer. 2013. "The Most Dangerous Idea in Federal Reserve History: Monetary Policy Doesn't Matter." *American Economic Review* 103, no. 3: 55–60.

15985-05b-Reis_Comm&GD-4thPgs.indd 376 5/12/22 12:07 PM

Yellen, Janet L. 2016. "Macroeconomic Research after the Crisis." Speech at The Elusive "Great" Recovery: Causes and Implications for Future Business Cycle Dynamics, 60th Annual Economic Conference sponsored by the Federal Reserve Bank of Boston, Boston, Massachusetts, October 14. https://www.federalreserve.gov/newsevents/speech/yellen20161014a.htm.

GENERAL DISCUSSION Robert Hall commented that extending the analysis from a concern of expectations—the first moment of the conditional distribution of future inflation—to the entire distribution is a big step forward. Hall also notes that Ricardo Reis does not say anything about current inflation. Reis noted that monetary policy response to changes in expectations may make inflation persistently higher, and he was exploring whether the whole distribution is a leading indicator for changes in the inflation regime.

Olivier Blanchard noted that the emphasis on expected inflation may be excessive in the current context. A central issue is whether workers who have seen prices increase and their purchasing power eroded will want to catch up and ask for higher wages. This is quite independent of their expectations about future inflation. If they do get increases in nominal wages, which are easier to get in a tight labor market, this might start a wage price spiral, give more momentum to inflation, and make it harder to decrease later.

Laurence Meyer considered what happened in 1968–1974 and suggested that this tells us nothing about what's relevant today; rather, it's an example of bad theory and bad policy, to which Reis responded that his point in the paper is that the bad policies were partly a result of ignoring inflation expectations data, missing their drift during that time. While this may or may not be where we are currently, it is a valuable lesson still today.

Meyer wondered why Reis does not mention the Index of Common Inflation Expectations (CIE), a weighted average of market-based and survey measures, which is a better measure of inflation expectations.[1] Meyer argued that, in contrast to what Reis suggests, the data show that inflation expectations are remarkably stable, as one would expect, and that we should not pay this kind of attention to quasi-supply shocks, although he agreed that there are many possible outcomes for inflation going forward and that supply shocks make inflation expectations difficult to predict.

1. See Hie Joo Ahn and Chad Fulton, "Index of Common Inflation Expectations," FEDS Notes, September 2, 2020 (Washington: Board of Governors of the Federal Reserve System), https://www.federalreserve.gov/econres/notes/feds-notes/index-of-common-inflation-expectations-20200902.htm.

Jason Furman then pondered about how the work of Reis compares to the Federal Reserve's CIE, which, he said, does not have any logic in how it relates to inflation expectations but only an internal logic of a set of correlates and, he suggested, is very inertial in its construction as an AR(4), with a slow effect of inflation expectations on the index. Furman wondered about where there's a structural break in these measures.

In response, Reis said that he finds the CIE to be a flawed measure, extracting a common component from a factor model that mixes indicators of fundamentally different variables: it has expectations from different people, markets, moments, and maturities, while ignoring the economics that tightly links them. Moreover, the factor model uses a sample of data to estimate correlations between these variables that likely changed dramatically in the last year, ultimately making it uninformative, Reis concluded.

Gerald Cohen recalled the substantial decline in inflation expectations in the mid-2010s, which did not see inflation decline, and asked whether Reis thinks that there is an asymmetry in the impact of inflation expectations or if this episode was just an outlier. Reis responded by referring to his paper where he finds a drift down in inflation expectations between 2012 and 2018 of 20 basis points—and even more in the eurozone. Significant and interesting, but much smaller than the drift up in the late 1960s (or today).[2]

Robert Gordon pointed out that the oil shock and the end of price control in the 1970s were both permanent in contrast to the current price increases, which will likely come down and add a negative component to inflation. The wide wage indexation during the 1970s, which translated into wage behavior and back into inflation in a vicious circle, is not an issue today, he continued, noting that the labor shortages and rising wages for low-skill workers that we see currently will likely only have level effects.

Reis agreed with Gordon but pointed out that he does not consider the period after the oil shocks in his paper. His claim is that the drift in expectations happened before, after 1967, and this earlier unanchoring is part of the reason why the oil shocks then had such a persistent and large effect on inflation. Reis admitted that we are still pretty far, though, from making accurate short-term inflation predictions using only inflation expectations data. Connecting Gordon's comment on price controls to the presentations of Yuriy Gorodnichenko and Alan Blinder, Reis said that he hopes future research will explore the disproportionate impact that price controls seem to have on inflation expectations.

2. Ricardo Reis, "The People versus the Markets: A Parsimonious Model of Inflation Expectations," discussion paper 15624 (London: Centre for Economic Policy Research, 2021).

Going forward, Frederic Mishkin worried about a policy mistake like that seen in the 1970s in light of the Federal Reserve's new average inflation targeting framework which, he argued, requires a level of transparency about the horizon of this average that the Federal Reserve has not provided—which risks not anchoring inflation expectations at all.[3] Consequences for demand are particularly important, he claimed, and he pondered how temporary the current inflation surge is. He summed up by saying that this is a critical juncture and that future policy needs to be laid out very carefully to not exacerbate inflation expectations and inflation, resulting in substantial tightening as the Federal Reserve responds.

To Mishkin's point about policy mistakes, Reis remarked that there is one error that he hopes the Federal Reserve won't make, and that his paper will modestly help: not looking at the expectations data properly.

3. Enrique Martínez-Garcia, Jarod Coulter, and Valerie Grossman, "Fed's New Inflation Targeting Policy Seeks to Maintain Well-Anchored Expectations," Federal Reserve Bank of Dallas, April 6, 2021, https://www.dallasfed.org/research/economics/2021/0406.

15985-05b-Reis_Comm&GD-4thPgs.indd 380 5/12/22 12:07 PM

DAVID AUTOR
Massachusetts Institute of Technology

DAVID DORN
University of Zurich

GORDON HANSON
Harvard Kennedy School

On the Persistence of the China Shock

ABSTRACT We evaluate the duration of the China trade shock and its impact on a wide range of outcomes over the period from 2000 to 2019. The shock plateaued in 2010, enabling analysis of its effects for nearly a decade past its culmination. Adverse impacts of import competition on manufacturing employment, overall employment-population ratios, and income per capita in more trade-exposed US commuting zones are present out to 2019. Over the full study period, greater import competition implies a reduction in the manufacturing employment-population ratio of 1.54 percentage points, which is 55 percent of the observed change in the value, and the absorption of 86 percent of this net job loss via a corresponding decrease in the overall employment rate. Reductions in population head counts, which indicate net out-migration, register only for foreign-born workers and the native born who are 25–39 years old, implying that exit from work is a primary means of adjustment to trade-induced contractions in labor demand. More negatively affected regions see modest increases in the uptake of government transfers, but these transfers primarily take the form of Social Security and Medicare benefits. Adverse outcomes are more acute in regions that initially had fewer college-educated workers and were more industrially specialized. Impacts are qualitatively—but not quantitatively—similar to those caused by the decline of employment in coal production since

Conflict of Interest Disclosure: The authors did not receive financial support from any firm or person for this article or from any firm or person with a financial or political interest in this article. David Autor is on the Panel of Economic Advisers for the Congressional Budget Office, is a trustee of the Urban Institute, and is a member of the advisory boards of the Opportunity and Inclusive Growth Institute at the Federal Reserve Bank of Minneapolis, the Malcolm Wiener Center for Social Policy at the Harvard Kennedy School, and WorkRise at the Urban Institute. Other than the aforementioned, they are currently not officers, directors, or board members of any organization with an interest in this article.

15985-06a_Autor-5thPgs.indd 381 5/12/22 12:06 PM

the 1980s, indicating that the China trade shock holds lessons for other episodes of localized job loss. Import competition from China induced changes in income per capita across local labor markets that are much larger than the spatial heterogeneity of income effects predicted by standard quantitative trade models. Even using higher-end estimates of the consumer benefits of rising trade with China, a substantial fraction of commuting zones appears to have suffered absolute declines in average real incomes.

Recent literature analyzing how import competition affects local labor markets has transformed our understanding of economic adjustment to globalization. Since at least David Ricardo, economists have appreciated that freer trade creates both winners and losers.[1] Yet, until recently, the prevailing view among scholars was that losses from trade were likely to be diffuse. The logic behind this view is simple and appealing: if labor is mobile across industries and regions, then the effects of greater import competition on, say, a labor-intensive and spatially agglomerated industry such as furniture making would be transmitted across all industries and regions employing less-educated labor.[2] Even though furniture factories in the North Hickory, North Carolina, commuting zone (CZ) may shut their doors, the pain felt by former furniture workers in the region would be little different from the losses suffered by those in equivalently skilled service sector jobs in Houston or Miami. We have since learned that losses from trade are often highly regionally concentrated.[3]

Much of the research that elucidates how local economies adjust to globalization is based on the China trade shock. In the early 1990s, China had both a large economy and, for its size, a narrow comparative advantage in manufacturing. Its rapid opening to trade and investment disrupted global markets. Economies producing the labor-intensive goods that China began exporting experienced a massive increase in global competitive

1. That Ricardo recognized winners and losers from trade is evident in his opposition to the British Corn Laws (Ricardo 1815), which he saw as redistributing income to landlords from workers (Maneschi 2008).

2. Labor mobility across industries and regions is core to the Heckscher-Ohlin trade model, for example (Feenstra 2016).

3. The seminal work of Jacobson, LaLonde, and Sullivan (1993) found scarring effects from job loss due to mass layoffs in the form of lower long-run earnings, which are more severe when the job loss occurs during a recession (Davis and von Wachter 2011). Although earlier work in trade found that exposure to imports can induce job loss (Grossman 1982; Kletzer 2001), the literature has only recently connected trade shocks to scarring effects from worker displacement.

15985-06a_Autor-5thPgs.indd 382 5/12/22 12:06 PM

supply, while those producing the raw materials and other inputs demanded by China felt a commensurately large increase in global demand. Recent work exploits the speed, scale, and concentrated impacts of China's reform-driven export growth to analyze the labor market consequences of globalization. Labor markets more exposed to import competition from China experienced more plant closures; larger declines in manufacturing employment, employment-population ratios, earnings for low-wage workers, housing prices, and tax revenues; and larger increases in childhood and adult poverty, single parenthood, and mortality related to drug and alcohol abuse, as well as greater uptake of government transfers.[4]

In this paper, we reexamine the local labor market impacts of exposure to Chinese import competition in order to evaluate which market adjustment mechanisms are operative and over what time horizons, how long adverse impacts endure, and how adjustment to trade compares to adjustment to other localized shocks. Among the most striking findings in recent work is that regions experiencing larger trade-induced reductions in manufacturing employment saw neither differential reductions in labor supply—due, for example, to out-migration—nor greater absorption of workers by non-manufacturing sectors. Manufacturing job loss translated nearly one for one into declines in the employment-population ratio.[5] Because in equilibrium changes in the employment rate are approximately proportional to changes

4. Research addresses regional and industry employment (Autor, Dorn, and Hanson 2013a; Pierce and Schott 2016; Handley and Limão 2017); plant closure (Bernard, Jensen, and Schott 2006; Acemoglu and others 2016; Asquith and others 2019); labor earnings (Autor and others 2014; Chetverikov, Larsen, and Palmer 2016); government transfers (Autor, Dorn, and Hanson 2013a; Autor and others 2014); housing prices and tax revenues (Feler and Senses 2017); migration (Autor, Dorn, and Hanson 2013a; Monte, Redding, and Rossi-Hansberg 2018; Faber, Sarto, and Tabellini 2019; Greenland, Lopresti, and McHenry 2019); and marriage, fertility, and mortality (Autor, Dorn, and Hanson 2019; Pierce and Schott 2020). Other work explores impacts on consumer prices (Bai and Stumpner 2019; Jaravel and Sager 2019; Hottman and Monarch 2020; Borusyak and Jaravel 2021), US exports (Feenstra, Ma, and Xu 2019), domestic innovation (Hombert and Matray 2018; Autor and others 2020b), and schooling (Greenland and Lopresti 2016). Additional work studies how trade shock impacts are mediated by labor market conditions (Austin, Glaeser, and Summers 2018), industry product cycles (Eriksson and others 2019), labor regulations (Chan 2019), and trade deficits (Dix-Carneiro and Kovak 2017). Autor, Dorn, and Hanson (2016) and Redding (2020) review the literature; and Dorn and Levell (2021) provide a comparative European-US perspective.

5. The absence of a local labor supply response to the China trade shock is consistent with findings on the null impact of generic Bartik (1991) labor demand shocks on local population head counts for less skilled workers (Bound and Holzer 2000; Diamond 2016; Notowidigdo 2020). Charles, Hurst, and Schwartz (2019) report comparable downward impacts on local employment rates from regional exposure to the national decline in manufacturing over 2000–2017.

15985-06a_Autor-5thPgs.indd 383 5/12/22 12:06 PM

in real income (Amior and Manning 2018; Galle, Rodríguez-Clare, and Yi 2017), trade-induced increases in joblessness may indicate adverse changes in welfare. Motivated by this concern, we explore the duration of shock impacts through 2019.

In more trade-exposed regions, we find that negative effects on manufacturing employment continue to build well beyond the culmination of the trade shock itself. Adverse impacts persist out to 2019, nearly a decade after the shock reached peak intensity. When examining the full study period, the midpoint for estimates using alternative data sources indicates that greater import competition implied a reduction in the manufacturing employment-population ratio of 1.54 percentage points, which is 55 percent of the observed change in the value, and that 86 percent of this manufacturing employment decline was absorbed by a decrease in the overall employment rate. Although the trade shock generated modest net out-migration, this occurred only among native-born adults age 25–39 and the foreign-born. Impacts of the trade shock thus appear to be long-lasting and to entail reduced participation in work, as evident in declines in wages and salaries and personal income in trade-exposed regions. Income losses are only minimally offset by the increased uptake of government transfers, which also remain elevated to the end of the analytic window.

Why has the China trade shock had such enduring consequences? One logical explanation is that the shock itself never stopped intensifying. The data do not support this conjecture, however. China's spectacular manufacturing export growth slowed dramatically in the last decade, by which point China's reform-driven boom had run its course and the government had begun to roll back some reforms (Lardy 2019; Brandt and others 2020; Brandt and Lim 2020). Nor did export growth simply move en masse from China to nearby countries. Whether we look at the US market presence of China alone or combined with Vietnam and other Southeast Asian nations, to which China has begun to offshore some labor-intensive activities, the trade shock reached peak intensity around 2010 and stabilized thereafter (Hanson 2020). Thus, the China trade shock did not unwind, but it did plateau, allowing us to examine its consequences for nearly a decade beyond its full expression.

We conduct preliminary analyses of two other explanations for why trade-exposed labor markets suffered long-lasting hardship. One hypothesis is that many traditional manufacturing regions were poorly positioned to recover from job loss because of a dearth of college-educated workers, who are in high demand by sectors that are expanding nationally (Glaeser, Scheinkman, and Shleifer 1995; Diamond 2016; Bloom and others 2019).

15985-06a_Autor-5thPgs.indd 384 5/12/22 12:06 PM

A second is that specialization in a narrow set of industries left these regions exposed to industry-specific shocks that, once industry decline was initiated, would set in motion a process that is self-reinforcing (Dix-Carneiro and Kovak 2017). We find some support for both the dearth of human capital hypothesis and the reverse agglomeration hypothesis. We also discuss other mechanisms that may be at work, though we do not provide a definitive, monocausal explanation.

To put the China trade shock in context, we examine adjustment to the secular decline of the coal industry and the impact of the Great Recession, the intensity of each of which varied sharply across regions. Regions more specialized in coal saw larger declines in employment rates, which remained suppressed more than a quarter century after the industry downturn began in 1980 (Black, Daniel, and Sanders 2002; Black, McKinnish, and Sanders 2005). Similarly, localities harder hit by the Great Recession had lower employment-population ratios ten years after the recession had officially ended, extending the findings in Yagan (2019). Although the long-run impacts of import competition from China do not appear to be unique qualitatively, the quantitative magnitudes of these impacts stand out relative to other shocks.

Much of the empirical trade literature estimates trade shock impacts by comparing changes in outcomes in more and less trade-exposed regions. This strategy identifies the relative impact of the trade shock—for example, whether manufacturing employment fell by more in more-trade-exposed locations—but not the absolute impact of the shock, for example, whether the trade shock reduced manufacturing employment nationally (Heckman, Lochner, and Taber 1998; Helpman 2018). Although relative impacts are informative about the distributional consequences of trade, they may not be indicative of changes in aggregate outcomes.[6] To evaluate the welfare impacts of trade, recent literature uses reduced-form empirical results to discipline the calibration of quantitative trade models. These calibrations suggest that US aggregate welfare gains from China's market opening were positive but small and that some *relatively* adversely affected regions may have suffered *absolute* welfare declines.[7] To investigate these effects further, we compare regional variation in real income changes in quantitative analyses with that implied by our reduced-form work, which imposes no

6. This issue also arises in empirical work in macroeconomics (Chodorow-Reich 2019, 2020; Guren and others 2021; Wolf 2019).

7. See Caliendo, Dvorkin, and Parro (2019); Rodríguez-Clare, Ulate, and Vásquez (2020); Adão, Arkolakis, and Esposito (2019); Galle, Rodríguez-Clare, and Yi (2017); Kim and Vogel (2020).

15985-06a_Autor-5thPgs.indd 385 5/12/22 12:06 PM

model structure. Standard quantitative models generate regional variation in income changes that is only one-quarter of what we find in reduced-form estimation. Even higher-end estimates of the cost-reducing impact of trade with China on US consumer prices (Jaravel and Sager 2019; Borusyak and Jaravel 2021) imply that a substantial number of regional economies suffered absolute declines in real incomes.

A final consideration regards the normative implications of adverse impacts of the China trade shock on labor markets in the United States and other countries. Few economists would interpret our empirical results as justifying greater trade protection. As expected, quantitative models indicate that US aggregate gains from trade with China are positive. Yet, the fact that the losses from trade are regionally concentrated and long-lasting suggests that existing policies failed to insulate workers from the disruptive impacts of globalization. We close the paper by discussing what we know, and do not know, about policies that help shield workers from concentrated job loss.

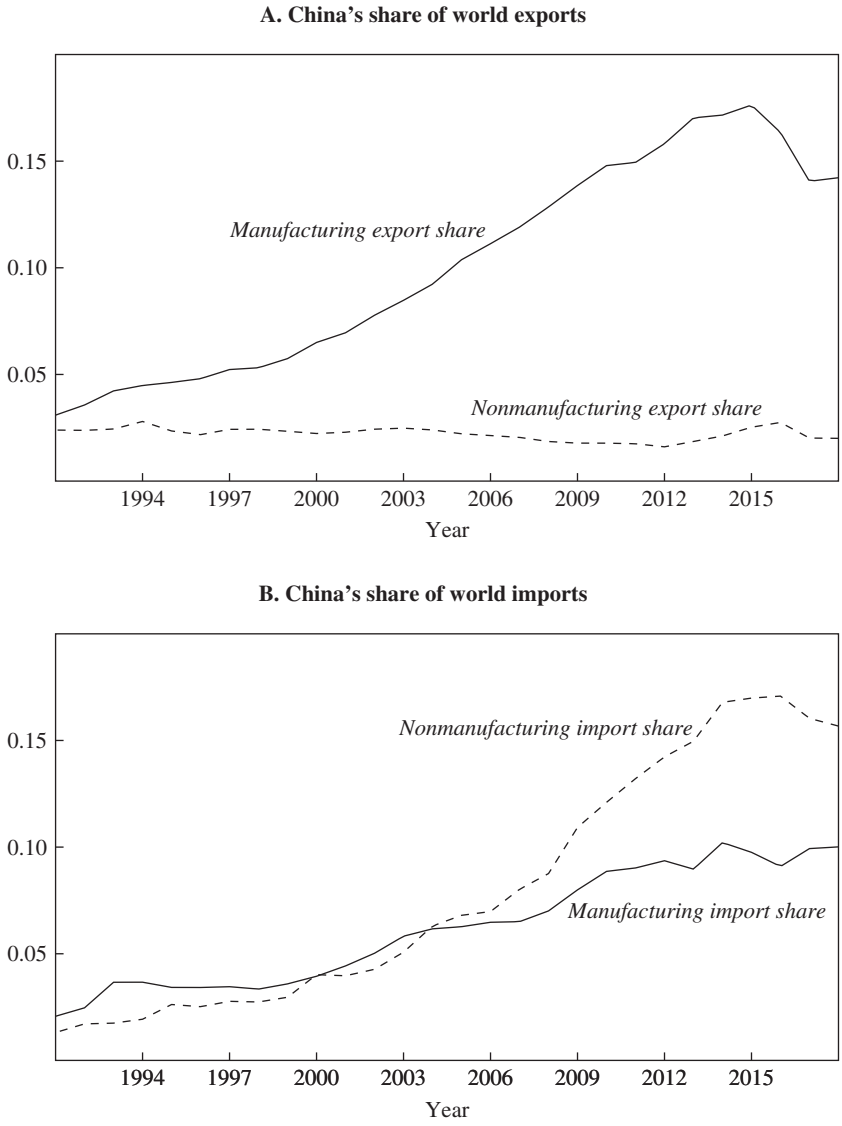
I. The Timing of the China Shock

We begin by reviewing China's recent export performance. Enduring impacts of the China trade shock on US local labor markets could be the result of the shock continuing to build after the 2000s and into the 2010s. We find this not to be the case. China's export growth has progressed through well-defined periods of initiation, acceleration, and stabilization. The country's share of the US market approached peak intensity around 2010 and plateaued thereafter.

That China's export growth has been explosive is universally understood. Its share of world manufacturing exports rose from 3.1 percent in 1991 to 17.6 percent in 2015, before dipping to 14.2 percent in 2018, as documented in figure 1, panel A.[8] This export boom has been concentrated in manufacturing and is recognized as transforming China into the world's factory.[9] Conversely, China's share of world exports of nonmanufactured goods has changed little over time, averaging just 2.3 percent between 1991 and 2018 and showing no discernible trend. This divergence in export growth between major sectors reflects China's powerful comparative advantage in manufacturing—particularly in labor-intensive manufacturing—which derived from the country's initial abundant supply of labor relative to its

8. Online appendix A.2 provides details on the data we use in this section.
9. See, for example, *Economist* (2015).

Figure 1. China's Share of World Trade by Major Sector



Source: UN Comtrade, SITC Revision 2.

supplies of agricultural land, natural resources, and physical capital (Wood 1995, 2018). Once China began to relax restrictions on foreign trade and investment, first through the expansion of special enterprise zones initiated by Deng Xiaoping (Naughton 2006), and later through an economy-wide liberalization tied to the country's accession to the World Trade Organization (WTO) in 2001 (Lardy 2019), it was able to realize this latent prowess and integrate into global value chains. The surge in manufacturing exports generated a corresponding increase in the import of raw materials used in industrial production, which mushroomed China's share of world nonmanufacturing imports from 1.2 percent in 1991 to 15.7 percent in 2018 (figure 1, panel B).[10]

Our concept of the China shock is the period spanning the country's reform-driven transition from an inward-oriented and heavily centrally planned economy to a comparatively open and market-oriented one in which the majority of production occurred in privately owned firms. We conventionally date the China trade shock as commencing in 1992, when Deng expanded his "reform and opening" agenda to include foreign trade and investment (Naughton 2006; Vogel 2011).[11] China's market transition unleashed dramatic improvements in productivity resulting from spatial and sectoral reallocation of resources (Hsieh and Klenow 2009; Brandt, Tombe, and Zhu 2013), the expansion of the private sector at the expense of inefficient state-owned enterprises (Song, Storesletten, and Zilibotti 2011; Khandelwal, Schott, and Wei 2013; Hsieh and Song 2015; Bai, Krishna, and Ma 2017), inflows of foreign capital and imported intermediate inputs (Yu 2015; Brandt and Morrow 2017), a reorientation of manufacturing to export assembly plants producing at the behest of multinational firms (Feenstra and Hanson 2005; Brandt and others 2017), and the adoption of previously banned foreign technologies (Hu and Jefferson 2009; Wei, Xie, and Zhang 2017; Li 2020).[12] Over 1998–2007, output per worker in Chinese manufacturing grew at the stunning annual average rate of

10. China's share of world manufacturing imports also rose over time, from 2.1 percent in 1991 to 10 percent in 2018, an increase driven in part by the expansion of global value chains whose assembly operations were anchored in China (Koopman, Wang, and Wei 2012; Kee and Tang 2016).

11. China's manufacturing export growth in fact began in the 1980s (Brandt and others 2020). Given the country's very low initial export levels, however, this early growth did little to change China's share of global exports.

12. The undervaluing of China's yuan-dollar exchange rate in the early 2000s may have also contributed to the country's export boom (Bergsten and Gagnon 2017), though without creating an industry bias in this growth.

15985-06a_Autor-5thPgs.indd 388 5/12/22 12:06 PM

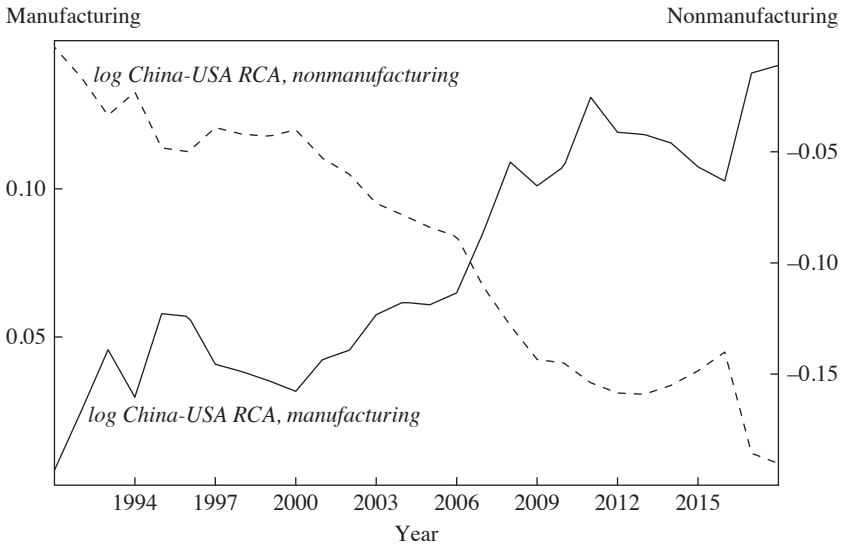
8.0 percent (Brandt, Van Biesebroeck, and Zhang 2012; Liu and Ma 2020). The 1992–2007 export surge thus appears to be an episode in which China's trade growth was driven largely by productivity gains resulting from its market opening.

Because much of China's post-Deng manufacturing growth was transitional, it was destined to be short-lived (Song, Storesletten, and Zilibotti 2011). Once a substantial share of the labor force had moved from the countryside to cities and from state-owned to private firms, easy growth may have been less attainable. Between 2010 and 2013, Brandt and Lim (2020) estimate that total factor productivity (TFP) growth in Chinese manufacturing was negative in eight of eleven subsectors, in stark contrast to earlier years. Changes in the political economy of reform may have further hampered manufacturing. In 2008, President Hu Jintao gave (low productivity) state-owned enterprises renewed prominence in industrial planning (Naughton 2016), a shift strongly reinforced by President Xi Jinping in 2012 (Lardy 2019). After 2007, entry of private and foreign-owned manufacturing firms fell sharply (Brandt and others 2020). The end of the productivity growth miracle and the roll back of pro-market reforms slowed manufacturing growth after 2010. Whereas the differential in annual manufacturing export growth between China and the rest of the world was 8.2 percentage points over 1991–2001 and 8.4 percentage points over 2001–2010, it was –0.4 percentage points over the 2010–2018 period.

China's post-2010 slowdown is also apparent when we compare its trade performance to the United States, which is helpful for defining the shock facing US manufacturing industries. A simple method to evaluate China's relative export success is to use the Balassa (1965) measure of revealed comparative advantage (RCA), which is the ratio of a country's share of world exports in a particular sector to its share of world exports of all goods.[13] In figure 2, we show the gap between the log China and log US RCA for manufacturing and nonmanufacturing. China began with a modest comparative advantage in manufacturing over the United States in 1991—as indicated by a China-US manufacturing versus all goods export differential of just 2.0 percentage points in that year. By 2011, its manufacturing versus all goods export differential over the United States had leaped to 13.1 percentage points, before falling and rising over the ensuing eight years. After 2010, the China-US relative RCA exhibits little trend.

13. In the Eaton and Kortum (2002) model of Ricardian comparative advantage, RCA summarizes differences in effective productivity between a country and the rest of the world. See also Hanson, Lind, and Muendler (2015).

Figure 2. China versus US Revealed Comparative Advantage



Source: UN Comtrade, SITC Revision 2.

Note: Revealed comparative advantage (RCA) is a country's share of world exports in a sector relative to a country's share of world exports of all merchandise.

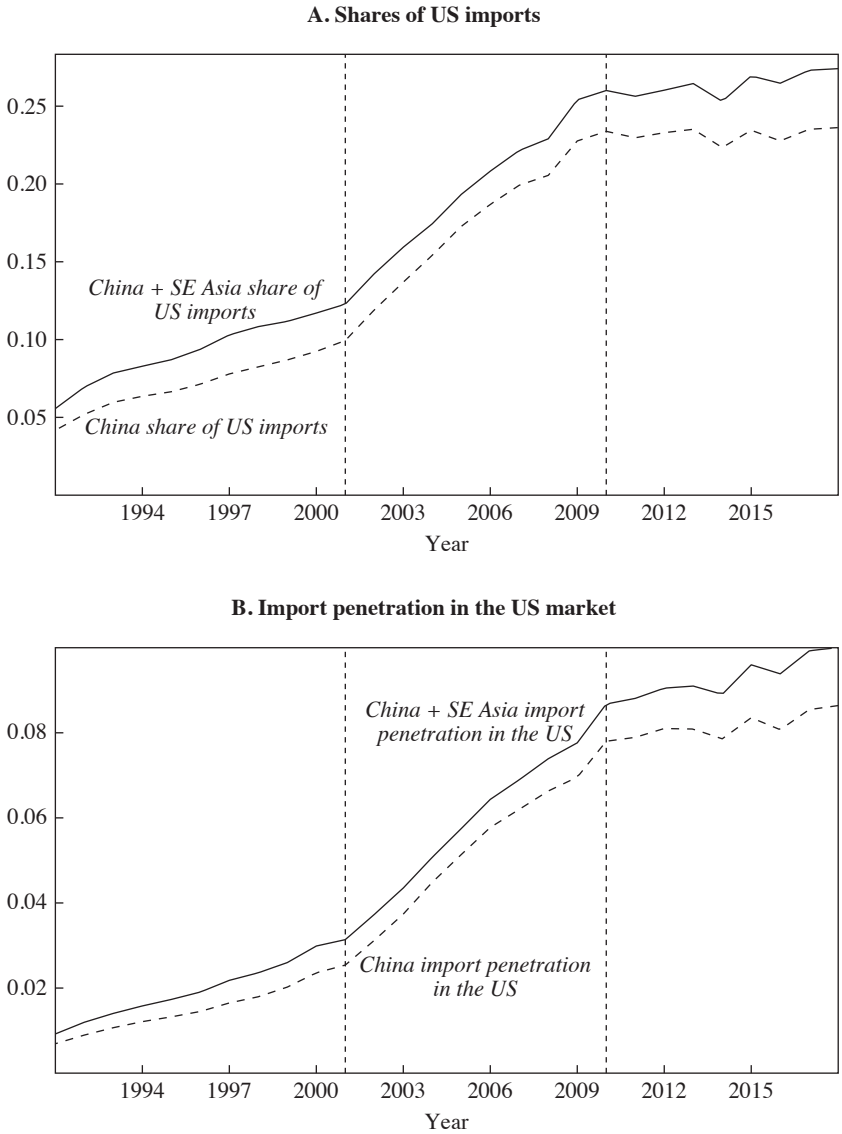
By construction, China's RCA in nonmanufacturing compared to the United States exhibits approximately the inverse pattern, dropping steadily through the 1990s and 2000s and then remaining at negative levels after 2011. China's comparative disadvantage in nonmanufacturing relative to the United States likely reflects the success of US firms in exporting agricultural products, certain mineral products, and oil and gas to the rest of the world, as well as market forces in China that have concentrated factors of production in manufacturing at the expense of other sectors.

Even though China's export growth in manufacturing slowed after 2010, it still could have contributed to overall US manufacturing import growth by offshoring production to other low-wage countries. Chinese firms have been active in building industrial parks for export production in Southeast Asia, and in Vietnam in particular.[14] In figure 3, we plot shares of US imports and import penetration in the US market for China alone and for China combined with low-income countries in Southeast Asia. We select these countries—Cambodia, Indonesia, Laos, Myanmar, Philippines,

14. See, for example, *Economist* (2016).

15985-06a_Autor-5thPgs.indd 390 5/12/22 12:06 PM

Figure 3. US Manufacturing Imports from China and Southeast Asia



Sources: UN Comtrade, SITC Revision 2; St. Louis Federal Reserve Bank; and authors' calculations.

Note: Import penetration in panel B is the ratio of US imports of manufactured goods to US domestic absorption (defined as gross output plus imports minus exports). All values exclude oil and gas.

15985-06a_Autor-5thPgs.indd 391 5/12/22 12:06 PM

Vietnam—based on their per capita GDP in 2010 being less than that of China.[15] We also add neighboring Bangladesh to this group because in recent decades multinational companies from Asia have expanded apparel factories in the country (Heath and Mobarak 2015). Figure 3 underscores that China's exports to the United States dwarf those of Southeast Asia. In 2010, China accounted for 23.4 percent of US manufacturing imports, whereas the Southeast Asian nations accounted for just 2.6 percent. In 2018, these figures were only modestly changed, at 23.6 percent and 3.8 percent, respectively. Although Southeast Asian countries did gain a larger share of US imports, the increase over the 2010 to 2018 period was just 1.2 percentage points. Import penetration, shown in figure 3, panel B, tells a similar story. Whether we look at China alone or combined with low-income Southeast Asian nations, its US market presence largely stabilized after 2010.[16]

Taken together, these trade patterns allow us to mark three distinct periods of China's export performance: the gradual initiation of China's export boom in the early 1990s, the dramatic acceleration of China's export growth following its WTO accession in 2001, and China's export plateau after 2010, which coincided with slowing manufacturing productivity growth. This slowdown may have been the result of China having completed its transitional period of post-reform growth, and possibly also of the partial reversal of reforms by Presidents Hu and Xi. Whatever the cause, the China trade shock appears to have stopped intensifying after 2010.

II. Estimating Labor Market Adjustment to Trade Shocks

This section presents our measure of the China trade shock and main empirical specification, which builds on Autor, Dorn, and Hanson (2013a) and Acemoglu and others (2016). We aim to identify the causal effect of import competition from China on labor market outcomes for US CZs, which is our measure of local labor markets (Tolbert and Sizer 1996; Dorn 2009), and evaluate impacts as the shock intensifies in the early 2000s, reaches peak intensity around 2010, and stabilizes thereafter.

15. In 2010, world development indicators show per capita GDP (at 2010 prices) of $4,560 in China and a range of $988 (Myanmar) to $2,130 (Philippines) in our Southeast Asian nations. The excluded, higher-income countries in the region are Brunei, Malaysia, Singapore, and Thailand, whose per capita GDPs range from $5,075 to $46,570; World Bank, "DataBank: World Development Indicators," https://databank.worldbank.org/source/world-development-indicators.

16. Results are qualitatively similar when we include additional low-income countries in South Asia (India, Pakistan, Sri Lanka), whose shares of global manufacturing exports have remained flat in recent years (Hanson 2020).

15985-06a_Autor-5thPgs.indd 392 5/12/22 12:06 PM

Table 1. Summary Statistics for CZ Import Penetration

	Change in CZ import penetration	*Mean (ppts)*	*Standard deviation*	*25th percentile*	*50th percentile*	*75th percentile*
Imports by United States	1991 to 2000	0.94	0.60	0.52	0.88	1.22
	2000 to 2007	1.35	0.84	0.84	1.15	1.62
	2000 to 2010	0.85	0.59	0.50	0.69	1.07
	2000 to 2012	0.89	0.59	0.51	0.75	1.17
	2000 to 2014	0.85	0.56	0.47	0.73	1.10
Imports by other high-income countries	1991 to 2000	0.70	0.48	0.40	0.63	0.87
	2000 to 2007	1.38	0.79	0.91	1.31	1.61
	2000 to 2010	1.27	0.82	0.82	1.13	1.47
	2000 to 2012	1.22	0.69	0.79	1.15	1.44
	2000 to 2014	1.01	0.58	0.66	0.96	1.21

Sources: UN Comtrade, SITC Revision 2; NBER-CES Manufacturing Industry Database; BEA Regional Economic Information System; National Vital Statistics System; and authors' calculations.

Note: This table shows the change in import penetration (100 × CZ weighted average change in imports from China/domestic industry absorption) for the 722 commuting zones in the continental United States and the instrumental variable for this value at given time horizons. Values are decadalized (× 10/length of time period) and weighted by the CZ working-age population in the initial year.

II.A. Baseline Specification

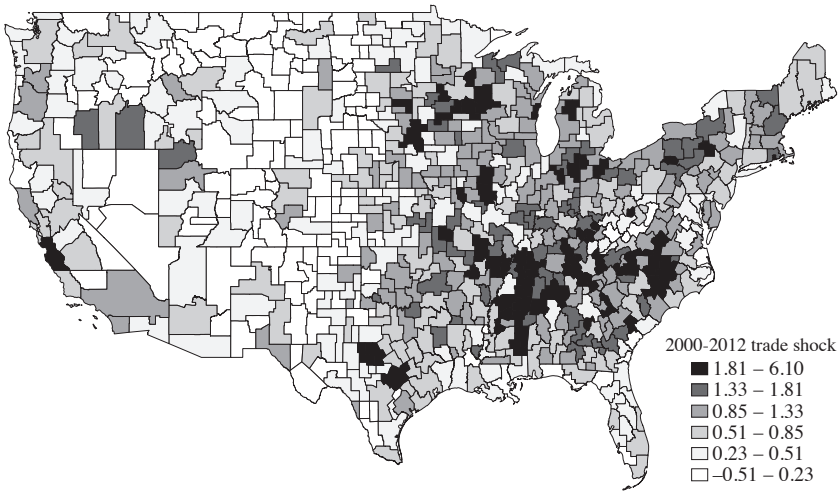
We examine exposure to import competition from China for the 722 CZs in the continental United States. As in Acemoglu and others (2016), our measure of trade exposure is the average change in Chinese import penetration across industries, weighted by industry shares in initial CZ employment:

$$\Delta IP^{cu}_{i\tau} = 100 \times \sum\nolimits_j s_{ijt} \Delta IP^{cu}_{j\tau}, \tag{1}$$

where $\Delta IP^{cu}_{j\tau} = \Delta M^{cu}_{j\tau}/(Y_{jt} + M_{jt} - X_{jt})$ is the growth of Chinese import penetration for US manufacturing industry j over time interval τ (2000 to 2012 in most cases), t is the base period (2000 in most cases), and $s_{ijt} \equiv L_{ijt}/L_{it}$ is the share of industry j in CZ i's total employment in the base year. We compute the change in import penetration as the growth in US industry imports from China, $\Delta M^{cu}_{j\tau}$, divided by initial industry domestic absorption (US industry shipments plus net imports, $Y_{jt} + M_{jt} - X_{jt}$), as summarized in table 1.[17] Differences in $\Delta IP^{cu}_{i\tau}$ across CZs stem from variation in local industry

17. We measure imports using Harmonized System (HS) trade data from UN Comtrade, harmonized to four-digit Standard Industrial Classification (SIC) industries, and industry shipments using the National Bureau of Economic Research (NBER) manufacturing productivity database (Autor and others 2014).

Figure 4. Chinese Import Penetration (2000–2012) in US Commuting Zones



Sources: UN Comtrade, SITC Revision 2; NBER-CES Manufacturing Industry Database; BEA Regional Economic Information System; National Vital Statistics System; and authors' calculations.

Note: This figure shows the change in import penetration from China in equation (1) over 2000–2012. The legend indicates values for the bottom four quintiles and the top two deciles.

employment in the base year, which arises from differential specialization in manufacturing and in import-intensive industries within manufacturing.

In figure 4, we map the China trade shock in equation (1) across CZs for the 2000–2012 time period. The most affected CZs, shown as those in the top two deciles of increased import penetration, are concentrated in the eastern half of the United States, and especially in the Southeast (north of the Deep South) and Midwest (outside of large metropolitan areas). These CZs are where US manufacturing relocated as it moved out of major cities in the Northeast and northern Midwest in the middle of the twentieth century (Eriksson and others 2019).[18]

To quantify the evolving impact of the China trade shock on labor market outcomes, we estimate first-difference models using successively longer time differences. Our regressions have the form,

$$\Delta Y_{it+h} = \alpha_t + \beta_{1h}\Delta IP^{cu}_{it} + \mathrm{X}'_{it}\beta_2 + \varepsilon_{it+h}, \quad (2)$$

18. Autor, Dorn, and Hanson (2013b) show that there is little correlation between a CZ's exposure to Chinese import competition and exposure to routine task-replacing technological change.

where ΔY_{it+h} is the change in an outcome for CZ i between the initial year t and later year $t + h$ for $h = 1, \ldots, T$, such that we estimate separate regressions for each time difference between 2000–2001 and 2000–2019. We focus on the period 2000–2019, which comprises the decade of China's most intense export growth, and the subsequent decade when China's export growth leveled off, up to the year before the onset of the COVID-19 recession in 2020. Our baseline definition of the trade shock is the period 2000–2012, whose first year is one year prior to China's WTO entry and whose final year postdates the culmination of the trade shock in 2010 and the volatility in global trade following the 2008 global financial crisis.[19] For $h < 12$, we estimate shock impacts for outcome periods that are less than complete—that is, whose length is less than that of the trade shock itself—which reveals how long it takes for the full impact of the shock to manifest. For $h > 12$, we estimate shock impacts on outcome periods that extend beyond the culmination of China's export boom, which reveals whether shock impacts attenuate over longer time horizons. If, for instance, it takes time for workers displaced from manufacturing to find jobs in other sectors or to migrate elsewhere, we may not see full adjustment along these margins until well after the shock reaches full intensity.

The control vector X'_{it} contains time trends for US Census divisions and start-of-period CZ-level covariates: the manufacturing share of employment, which allows us to focus on trade exposure arising from the within-manufacturing industry mix; specialization in occupations according to their routine-task intensity and offshorability (based on Autor and Dorn 2013), thus accounting for exposure to automation and non-China-specific globalization; the fractions of foreign-born and non-white workers, the college-educated portion of the population, and the fraction of working-age women who are employed, which absorbs variation in outcomes related to labor-force composition; and the population shares of residents age 0–17, 18–39, and 40–64, which control for demographic factors that may affect labor force participation and eligibility for government transfers (see online appendix table A2). For all outcomes other than population head counts,

19. We show in online appendix figure A4 that we obtain nearly identical coefficient estimates when using trade shocks constructed for the 2000–2007, 2000–2010, 2000–2012, and 2000–2014 time periods. This similarity in results is a consequence of the post-2000 trade shock being front-loaded on the 2000–2007 period. The 2000–2007 shock is 79.8 percent of the magnitude of the 2000–2012 shock and 88.8 percent of the magnitude of the 2000–2014 shock. The CZ-level pairwise correlations of the 2000–2007, 2000–2010, 2000–2012, and 2000–2014 trade shocks range from 0.93 to 0.96.

15985-06a_Autor-5thPgs.indd 395 5/12/22 12:06 PM

regressions are weighted by the CZ working-age population (age 18–64) in the initial year; for head count regressions, weights are the total CZ population in the initial year. All specifications cluster standard errors at the state level.

We focus on US imports from China and not US exports to China because the former dwarf the latter and because our instrumentation strategy (see below) is less well-suited to isolating exogenous variation in US export growth.[20] Autor, Dorn, and Hanson (2013a) find similar results when replacing growth in US imports with growth in US net imports. Consistent with these findings, Feenstra, Ma, and Xu (2019) detect no impact of growth in US exports to China on US manufacturing employment, although overall US export growth is positively associated with manufacturing activity.[21]

In related work, Eriksson and others (2019) study the Japan trade shock of 1975–1985 (during which Japan's exports of autos and other durables surged) and the East-Asian Tiger trade shock of 1975–1988 (during which Hong Kong, Korea, Singapore, and Taiwan had manufacturing export booms), which occurred a dozen years before China's WTO accession. They detect no impact of the Japan shock on employment rates in US CZs, while Batistich and Bond (2019) find a similar overall null effect of the shock on CZ manufacturing employment, though a differentially negative effect on Black relative to white manufacturing workers. It is further the case that the CZs exposed to the earlier East Asian export surges had limited overlap with those exposed to the later China trade shock. The former group had relatively high levels of income, college attainment,

20. The 2000–2012 increase in US manufacturing imports from China ($292 billion) was 4.1 times the increase in US manufacturing exports to China ($71 billion), for values in 2015 US dollars.

21. Weak US manufacturing exports to China could be a result of US comparative advantage lying elsewhere, such as in agriculture, oil and gas, and intellectual property. Dorn and Levell (2021) show that most major European economies experienced a combination of rapidly increasing Chinese import penetration and modestly growing exports to China that was qualitatively comparable to the US experience, with the notable exception of Germany where manufacturing exports (dominated by machine tools and cars) grew in parallel to imports (dominated by labor-intensive products). In Germany, job loss in import-exposed regions counteracted job growth in export-oriented ones (Dauth, Findeisen, and Suedekum 2014). OECD countries in which exports to China grew to a similar extent as imports from China experienced smaller declines in manufacturing employment than countries like the United States, where imports strongly dominated exports (Dorn and Levell 2021). Weak US export growth could also be attributable to the forces behind large trade surpluses in China and trade deficits in the United States (Autor, Dorn, and Hanson 2016; Dix-Carneiro and others 2021).

15985-06a_Autor-5thPgs.indd 396 5/12/22 12:06 PM

and patents per capita, whereas the latter had relatively low values of these indicators.[22]

Import penetration in equation (1) includes all types of shipments from China to the United States, whether they emanate from Chinese companies exporting final goods to the US market on their own accord or export processing plants—such as Foxconn, which assembles Apple iPhones from imported parts and components—that are part of global value chains controlled by multinational enterprises. Boehm, Flaaen, and Pandalai-Nayar (2020) find that much of the post-1990 employment decline in US manufacturing occurred in establishments owned by US multinationals that were simultaneously expanding operations in their foreign subsidiaries. Between 1995 and 2005, export processing plants contracting with multinationals accounted for a hefty 55 percent of China's manufacturing exports (Feenstra and Hanson 2005). Over time, however, Chinese firms have expanded input production in China, which reduced the share of processing plants in China's exports to 35 percent by 2015 (Ma 2020). Unfortunately, our data do not allow us to separate US imports from China into those that came through global value chains versus other channels. In recent work, Aghion and others (2021) use firm-level data for France to separate imports from China into "horizontal" products—imports in product categories similar to a firm's exports—and "vertical" products—imports in product categories similar to a firm's imported inputs. Whereas exposure to the first type of imports predicts employment declines, exposure to the second type of imports does not. These results suggest that imports that intensify head-to-head competition with domestic firms spur contraction of domestic manufacturing.[23]

II.B. Causal Identification

A challenge for identifying the causal impact of import exposure on labor-market outcomes in equation (2) is that US imports may change both because of shocks to US product demand and shocks to foreign product

22. Despite a null net impact of the Japan shock on overall employment levels, Borjas and Ramey (1995) find that US metropolitan areas more exposed to import competition during the East Asian shock period had large increases in college-noncollege wage differentials. Batistich and Bond (2019) find further that the Japan shock caused changes in composition of employment in exposed CZs, with employment falling for Black workers and rising for white workers.

23. In Aghion and others (2021), adverse impacts of import exposure are concentrated on low-productivity firms. In the US market, by contrast, Bernard, Jensen, and Schott (2006) and Autor and others (2014) find that trade-induced employment declines are concentrated in larger manufacturing establishments, consistent with the framework in Holmes and Stevens (2014).

15985-06a_Autor-5thPgs.indd 397 5/12/22 12:06 PM

supply, where the former may be correlated with the residual, ε_{it+h}. To identify the foreign supply–driven component of US imports from China, we follow Autor, Dorn, and Hanson (2013a) and Acemoglu and others (2016) in instrumenting US import exposure, $\Delta IP^{cu}_{i\tau}$, using non-US exposure, $\Delta IP^{co}_{i\tau}$, which we measure as the industry-level growth of Chinese exports to eight other developed countries:[24]

$$\Delta IP^{co}_{i\tau} = \sum_j s_{ijt-10} \Delta IP^{co}_{j\tau}, \tag{3}$$

where $\Delta IP^{co}_{j\tau} = \Delta M^{co}_{j\tau}/(Y_{jt-3} + M_{jt-3} - X_{jt-3})$. This expression differs from equation (1) by using imports from China in other high-income markets ($\Delta M^{co}_{j\tau}$) in place of US imports ($\Delta M^{cu}_{j\tau}$), the three-year lag of industry absorption ($Y_{jt-3} + M_{jt-3} - X_{jt-3}$) in place of its base year t value, and the ten-year lag of CZ industry employment shares, $s_{ijt-10} \equiv L_{ijt-10}/L_{it-10}$, in place of year t values (see table 1).[25]

Analyses of the China trade shock have used $\Delta IP^{co}_{i\tau}$ as a shift-share instrument in local labor market regressions and have directly applied the industry-level shocks, $\Delta IP^{co}_{j\tau}$, as instruments for the growth in industry-level import penetration (Autor and others 2014; Autor and others 2020b). Recent literature formalizes the basis for identification and inference in such shift-share settings. Borusyak, Hull, and Jaravel (2022) treat identification as based on exogeneity of the shifts—that is, the industry-level changes in import penetration—which clarifies the identifying assumptions in the analysis and provides an asymptotic rationale for using the shift-share structure to construct standard errors.[26] Adão, Kolesár, and Morales (2019) present a related method for estimating standard errors. Although these refinements of the shift-share approach do not change the fundamental conclusions about the impact of the China trade shock on regional employment, they do formalize exclusion restrictions and advance methods for

24. The eight comparison countries—determined by the availability of comparable HS trade data for the full sample period—are Australia, Denmark, Finland, Germany, Japan, New Zealand, Spain, and Switzerland.

25. The use of lagged values helps reduce both the role of simultaneity and the influence of measurement error.

26. The assumption of exogenous shifts is evident in industry-level analyses of the China shock that instrument a US industry's import growth from China with the industry's Chinese imports in other countries (Autor and others 2014; Acemoglu and others 2016). Goldsmith-Pinkham, Sorkin, and Swift (2020) develop an alternative approach to shift-share analysis, which is based on exogeneity of the shares (i.e., preexisting industry structures). As we discuss in online appendix A.4, their approach appears to be less suitable for our application. See Ferman (2019) for further discussion.

15985-06a_Autor-5thPgs.indd 398 5/12/22 12:06 PM

statistical inference. In online appendix A.4, we draw on this recent literature to discuss the basis for econometric identification, present results using alternative methods for constructing standard errors, and provide analysis of pre-trends in the data.

III. Empirical Results

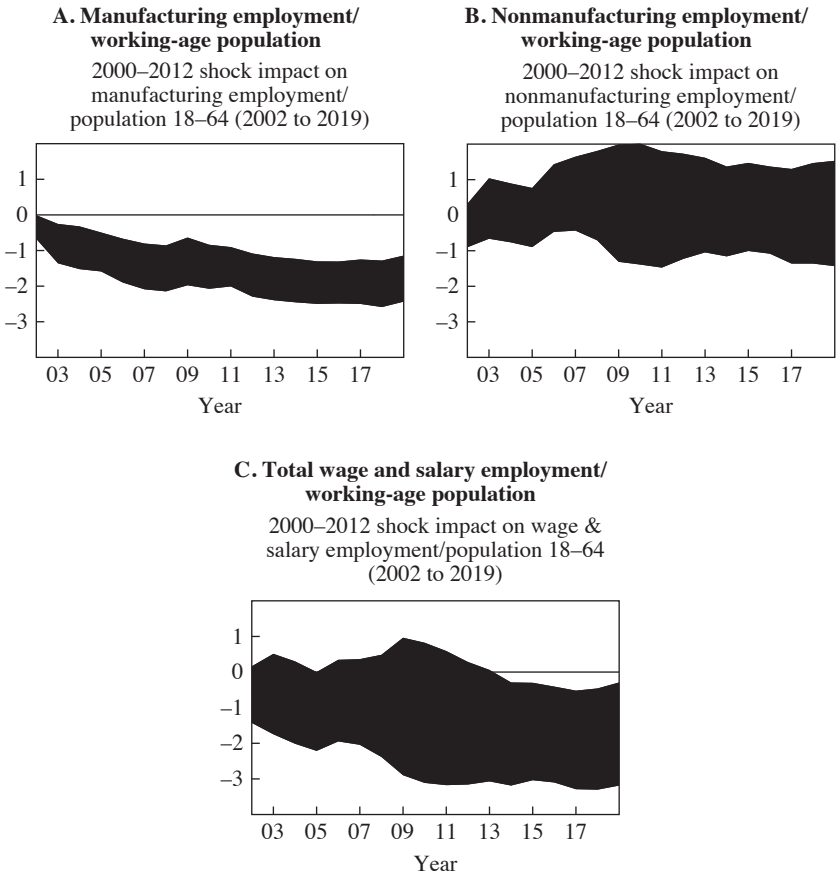
This section presents our main empirical results for the impact of trade exposure on US CZs. We use equation (2) to estimate how the 2000–2012 trade shock affected CZs over 2000–2019, focusing on three sets of outcomes. The first are ratios of manufacturing employment, nonmanufacturing employment, and total wage and salary employment to the working-age population (defined as individuals age 18–64). We measure employment using the Bureau of Economic Analysis (BEA) Regional Economic Information System (REIS) and population using the National Vital Statistics System, where we aggregate data from the county to the CZ level. Trade shock impacts on these outcomes reveal the direct consequences on manufacturing employment, and the magnitude of labor reallocations to nonemployment and other sectors. A second set of outcomes is population head counts, overall and by nativity and age-based subgroups, which reveal whether exposure to import competition led to net reductions in the resident population (e.g., due to out-migration). A third set of outcomes is per capita personal income, labor compensation, and government transfers, overall and by program type, which reveal the impact of trade shocks on average income and its components. We measure income, earnings, and government transfers using county-level data from the REIS. For most series, we use 2000 as the initial year.[27] Summary statistics on outcome and control variables are reported in the online appendix tables A1 and A2; online appendix A.2 contains further details on data sources and variable construction.

III.A. Employment Outcomes

We begin with the impact of the China trade shock on the CZ employment-population ratio, where we measure employment either as the total number of wage and salary jobs in a CZ or by splitting these jobs into manufacturing and nonmanufacturing (where the latter is defined as total wage and salary employment less manufacturing employment). Figure 5 displays two-stage least squares estimates of impact coefficients for trade exposure from the

27. Because of BEA changes in industry classification from the SIC to the North American Industry Classification System (NAICS) in 2001, our REIS series for manufacturing and nonmanufacturing employment begins in that year, rather than 2000.

Figure 5. Trade Shock Impact on Employment, 2001–2019



Source: Authors' calculations.

Note: Panels report two-stage least squares coefficient estimates for β_{1h} in equation (2) and 95 percent confidence intervals for these estimates. The dependent variable is the change in the specified outcome between 2001 and the year indicated on the horizontal axis; the trade shock is the decadalized 2000–2012 change in CZ import exposure, as defined in equation (1) and instrumented by equation (3). Control variables include initial-period CZ employment composition (shares of employment in manufacturing, routine-task-intensive occupations, and offshorable occupations, as well as employment share among women), initial period CZ demographic conditions (shares of the college educated, the foreign born, non-white individuals, and those age 0–17, 18–39, and 40–64 in the population), and census region dummies. Regressions are weighted by the CZ working-age population in 2000; standard errors are clustered by state.

15985-06a_Autor-5thPgs.indd 400 5/12/22 12:06 PM

time difference specification in equation (2) when using equation (3) to instrument for equation (1). Each panel shows results for eighteen separate regressions, where the horizontal axis organizes specifications according to the time difference being considered, beginning with 2001–2002 and extending out to 2001–2019. Vertical bars show 95 percent confidence intervals for the coefficient estimates. Structured in this manner, figure 5 shows the cumulative impact of trade exposure for progressively longer time differences (one to eighteen years).

The results for manufacturing employment in figure 5, panel A, show a negative effect of trade exposure that builds quickly and then stabilizes. The impact coefficient is −1.45 (t-value = −4.41) for 2001–2007, which is the end year of analysis in Autor, Dorn, and Hanson (2013a), reaches −1.90 (t-value = −6.24) for the 2001–2015 period, and then attenuates modestly to −1.79 (t-value = −5.42) for the 2001–2019 period. As one moves past 2012, the negative effect of greater import competition on manufacturing employment continues to grow, even though the trade shock itself no longer appears to be intensifying. Given a decadalized increase in import penetration over 2000–2012 of 0.89 percentage points (see table 1), the implied reduction in the manufacturing employment share over 2001–2019 is −1.59 (= −1.79 × 0.89) percentage points, or 59.3 percent of the overall 2001–2019 change in the manufacturing employment share of −2.68 percentage points (online appendix table A1). Alternatively, when comparing CZs at the 25th and 75th percentiles of trade exposure, the latter would be predicted to have a reduction in its manufacturing employment share that is 1.18 (= −1.79 × [1.17 − .51]) percentage points larger than in the former over 2001–2019. This compares to the 25th–75th percentile differential change in the manufacturing employment share of −2.17 (= −3.79 + 1.62) percentage points for the same period. Exposure to import competition from China thus appears to account for a large share of net manufacturing job loss in US CZs after 2000.[28]

Those losing jobs in manufacturing may move into other sectors or out of employment. Simultaneously, they may move to another CZ.[29] In figure 5,

28. These sustained negative impacts of trade exposure on local manufacturing employment for the United States are consistent with an analysis by Dix-Carneiro and Kovak (2017) for Brazil where trade liberalization in the 1990s persistently reduced employment in exposed regions, with even larger magnitudes than China trade impacts for the United States.

29. Workers exiting manufacturing employment may also end up incarcerated, enlisted into military service, homeless, or deceased. We observe these outcomes imperfectly in the data, but they are likely to be small in aggregate relative to the margins that we observe. Workers exiting the labor force may be of course replaced by new entrants, offsetting these losses. Our analysis captures net effects of the trade shock on employment.

15985-06a_Autor-5thPgs.indd 401 5/12/22 12:06 PM

panel B, we consider the impact of exposure to import competition on the CZ share of the working-age population employed in nonmanufacturing industries. Impact coefficient estimates are close to zero at all time intervals and are imprecisely estimated (e.g., the impact coefficient for the 2001–2019 period is 0.05, *t*-value = 0.06). We see no evidence that nonmanufacturing sectors absorbed local workers released from manufacturing due to the China trade shock.[30] The wide standard error bounds in figure 5, panel B, are suggestive of heterogeneity across CZs in how nonmanufacturing absorbs displaced manufacturing workers, a possibility we explore in section III.D.

The fall in manufacturing employment and absence of offsetting job gains in nonmanufacturing imply a decline in total wage and salary employment in trade-exposed CZs, confirmed by figure 5, panel C. Decreases in the employment-population ratio are comparable in magnitude to those for manufacturing employment. The effect reaches its maximum value for the 2001–2017 period: the impact coefficient of −1.91 (*t*-value = −2.70) compares to −1.88 (*t*-value = −5.89) for manufacturing employment at the same horizon. The impact attenuates modestly late in the period, ending at −1.74 (*t*-value = −2.35) over 2001–2019, which is close to the manufacturing impact at this horizon. At long time horizons, much of the net absorption of manufacturing job loss is through increases in nonemployment. Amior and Manning (2018) argue that changes in the local employment-population ratio summarize changes in local average real income. By this logic, the China trade shock would have substantially altered the distribution of well-being across CZs, a possibility we explore in more detail below through our analysis of personal income per capita.

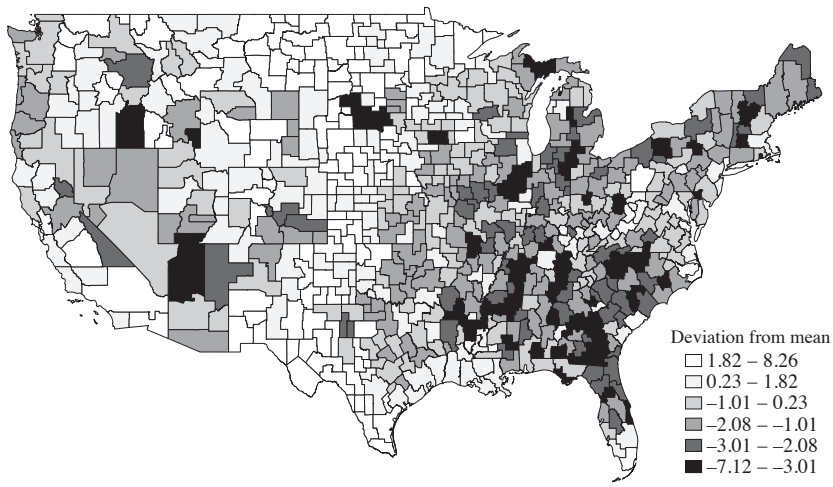
Figure 6, panel A, maps the actual change in wage and salary employment-population ratios across CZs for the 2001–2019 time period. While the national employment rate barely changed over this period (+0.19 percentage points), the map reveals considerable variation across space, with lower employment growth in parts of the South, Midwest, and Northeast, and faster employment growth in the Great Plains and some of the western and northeastern coastal areas. In figure 6, panel B, we map the implied impact of the China trade shock (as measured in equation (1) for 2000–2012) on changes in employment-population ratios for the same 2001–2019 time

30. Bloom and others (2019) report evidence that CZs experiencing greater import competition do see larger increases in nonmanufacturing employment. We do not find this to be the case. We do, however, confirm their finding that employment losses spurred by the China trade shock are highly durable in CZs with a less-educated population, while they are ultimately more than fully offset in CZs with a more-educated population (see section III.D).

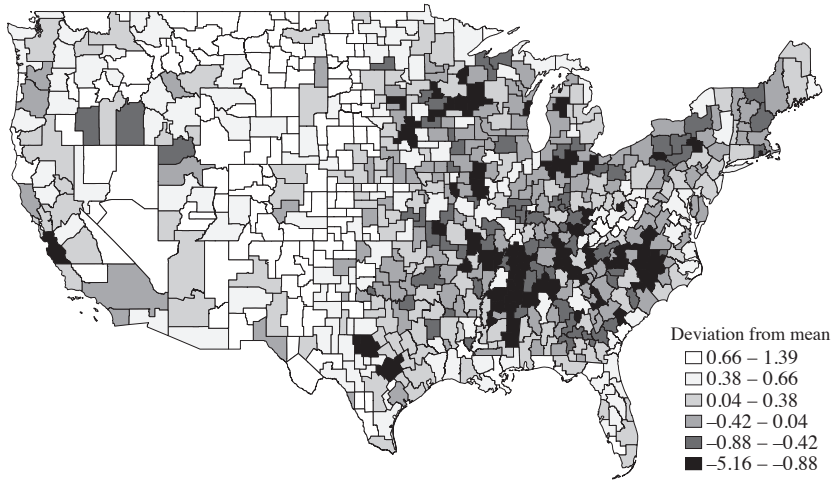
15985-06a_Autor-5thPgs.indd 402 5/12/22 12:06 PM

Figure 6. Implied Impact of Import Competition 2001–2019

A. Actual change in total employment-population ratio



B. Predicted change due to the China trade shock



Sources: UN Comtrade, SITC Revision 2; NBER-CES Manufacturing Industry Database; BEA Regional Economic Information System; National Vital Statistics System; and authors' calculations.

Note: Panel A plots the observed change in the ratio of wage and salary employment to working-age population over 2001–2019; panel B plots the implied change in this value due to the China trade shock based on our estimates (i.e., the product of the change in import penetration over 2000–2012 in equation (1) multiplied by the two-stage least squares coefficient estimate for this variable from the specification in equation (2) in which the dependent variable is the 2001–2019 change in the ratio of wage and salary employment to the working-age population, as shown in figure 4, panel C). Legends indicate values for the bottom four quintiles and the top two deciles of this value, arranged in order of shock intensity.

15985-06a_Autor-5thPgs.indd 403 5/12/22 12:06 PM

period. A visual comparison between the two panels in figure 6 shows a striking correlation, as many of the CZs that lost employment overall (in panel A) were also more adversely affected by the trade shock (in panel B). This correlation suggests that the China shock had an important influence on the differential employment growth across US regions in the last two decades. The top 5 percent of CZs in terms of implied reductions in employment-population ratios, which are listed in online appendix table A4, include a preponderance of locations that in 2000 were relatively highly specialized in manufacturing and had relatively few college-educated workers. In 2000, thirty-three of the thirty-eight most exposed CZs had a manufacturing employment share above 25 percent, relative to the national population-weighted median of 15.4 percent, and thirty-three of thirty-eight had a college-educated share of the working-age population below 20 percent, relative to the national population-weighted median of 23.4 percent. The distinctiveness of trade-impacted CZs motivates the heterogeneity analysis that we undertake in section III.D.

To test robustness, online appendix figure A5 replicates the analysis in figure 5, now using US Census and American Community Survey (ACS) data to construct changes in employment-population ratios.[31] Census-ACS data have the advantage over the REIS of measuring employment as the number of individuals with a job rather than the total number of jobs in a CZ. A disadvantage of these household survey–based data is that they offer comparatively small sample sizes in annual samples after 2000, requiring us to use combined ACS annual surveys. We estimate the impact of the China trade shock on employment-population ratios for three time differences: 2000 (using data from the 2000 census 5 percent sample) to 2007 (using pooled 1 percent annual samples from the 2006–2008 ACS files), 2012 (using pooled 1 percent annual samples from the 2011–2013 ACS files), and 2018 (using pooled 1 percent annual samples from the 2017–2019 files).[32] The 2000–2007 period matches the later time period studied in Autor, Dorn, and Hanson (2013a), the 2000–2012 period matches the duration of our trade shock measure, and 2000–2017/2019 is the longest post-2000 period that we can study using the census and ACS.

Focusing on results at the long horizon (i.e., 2000–2018 in census-ACS data, 2001–2019 in REIS data), we see that in figure A5 in the online

31. US Census Bureau, "American Community Survey (ACS)," https://www.census.gov/programs-surveys/acs.

32. See online appendix table A3 for summary statistics; other regression variables are the same as in figure 5.

15985-06a_Autor-5thPgs.indd 404 5/12/22 12:06 PM

appendix the impact coefficient for manufacturing employment-population of −1.67 (*t*-value = −7.00) compares to −1.79 (*t*-value = −5.42) in figure 5, panel A. Similarly, in figure A5 the impact coefficient for total employment-population of −1.23 (*t*-value = −3.32) compares to −1.74 (*t*-value = −2.35) in figure 5, panel C. These two sets of results appear quite comparable in light of the differences between the data sources, specifically: using census-ACS data-based estimates with 2000 as a start year (instead of 2001), combining 2017–2019 data for the 2019 observation, and measuring employment as persons employed instead of the total number of jobs. In census-ACS data, the implied trade-induced reduction in the manufacturing employment share is −1.49 (= −1.67 × 0.89) percentage points, or 50.9 percent of the overall 2000–2018 change in the manufacturing employment share of −2.92 percentage points (online appendix table A3); recall that when using REIS data, the corresponding figure is 59.3 percent. The two sets of results differ mildly in that the long-run impact of the China trade shock on the nonmanufacturing employment ratio is larger in figure A5, panel b ($\beta = 0.43$, *t*-value = 1.35) than in figure 5, panel B ($\beta = 0.05$, *t*-value = 0.06), though the former estimate is imprecise and too small to prevent a decline in the overall employment rate.[33] Thus, in census-ACS data the long-horizon, trade-induced decline in the overall employment-population ratio (given by the impact coefficient −1.23) absorbs 73.7 percent of the trade-induced decline in manufacturing employment (given by the impact coefficient −1.67), whereas in REIS data the corresponding figure is 97.2 percent (= 1.74 / 1.79).

One of the most surprising results reported by Autor, Dorn, and Hanson (2013a) is that the adverse impacts of the China shock on manufacturing and total employment-to-population persisted over at least a decade. At the time that paper was written, the China trade shock was ongoing, making it infeasible to distinguish short-term from steady-state (or medium-run) impacts. The evidence presented above makes clear that even in 2019, nine years after the plateau of the China trade shock, there is no recovery of manufacturing or total employment-to-population rates in trade-exposed CZs. To a first approximation, CZs that are more trade-exposed suffered a durable reduction in the size of their manufacturing labor force, with the bulk of this decline translating into a long-run increase in nonemployment. The consequences of these disruptions went far beyond earnings and

33. Because the China trade shock began in the 1990s, one may view the specification in equation (2) as incomplete in that it does not control for the previous decade's trade shock. Online appendix section A.5.1 explores these adjustment dynamics.

employment. Job loss engendered social dislocation, in the form of lower marriage rates, increased single parenthood, a higher incidence of children living in poverty, and increased mortality from drug and alcohol abuse, especially among young males (Autor, Dorn, and Hanson 2019; Pierce and Schott 2020).

III.B. Spillovers across Regions and Industries

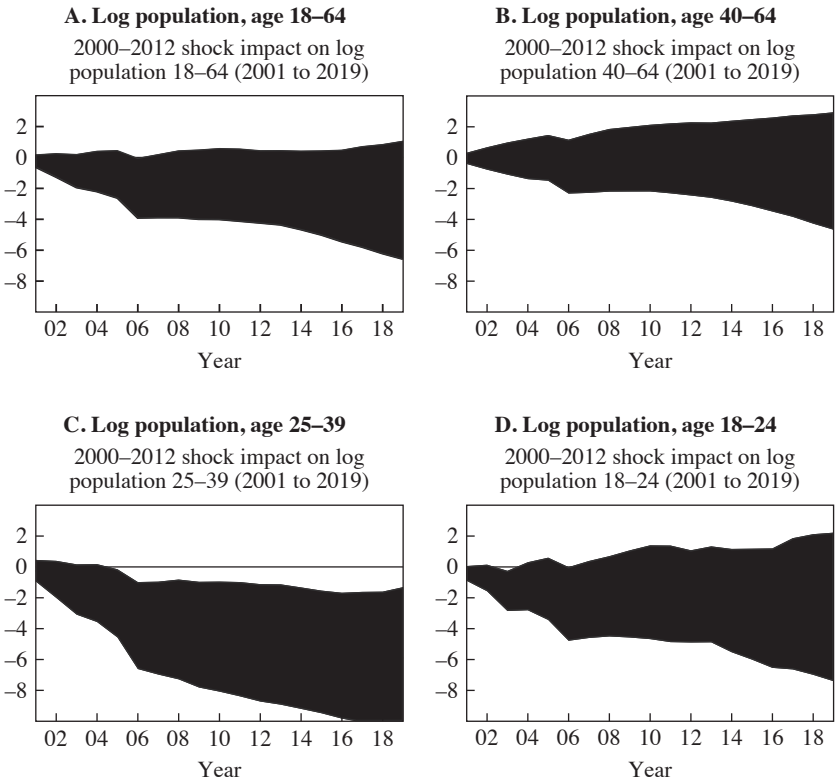
The empirical specification in equation (2) is consistent with a setting in which CZs are functionally small open economies. If changes in economic conditions in a given CZ do not materially affect outcomes in other CZs, we can examine each CZ on its own without modeling the transmission of shocks across locations. Of course, trade shocks having a direct impact on one CZ may be transmitted to other CZs via changes in regional flows of goods and production factors. We next evaluate evidence of cross-region and cross-industry spillovers of trade shocks.

CHANGES IN LOCAL LABOR SUPPLY The impact of import competition on CZ population head counts summarizes the net effect of trade shocks on the pool of both potential workers and nonworking residents (e.g., children, the elderly, and working-age nonparticipants). Labor supply responses to negative labor demand shocks may in turn differ by worker age. Younger workers, in particular, are likely to be relatively mobile (Bound and Holzer 2000). We examine the responsiveness of population head counts to greater import exposure separately for workers 18–24, 25–39, and 40–64 years old. This analysis is complicated by the fact that there are strong secular trends in population growth across US regions which began well before the China trade shock (Blanchard and Katz 1992). Greenland, Lopresti, and McHenry (2019) suggest that the approach used in Autor, Dorn, and Hanson (2013a) for evaluating the impact of trade shocks on labor supply, which conditions on a control vector similar to that in equation (2), does not account for such dynamics.[34] Accordingly, when we estimate equation (2) for log population head counts, we additionally include as a control the log change in CZ population over 1970–1990 in order to capture long-standing trends in population growth.

Figure 7 reports estimation results for population head counts. In figure 7, panel A, we find negative but insignificant impacts of trade exposure on the size of the working-age population, though the point estimates grow as

34. Estimation results for the employment-population ratio are not subject to this critique because using the ratio of employment to population effectively differences out population growth trends.

Figure 7. Trade Shock Impact on Log Population Head Counts, 2000–2019



Source: Authors' calculations.

Note: Panels report two-stage least squares coefficient estimates for β_{1h} in equation (2) and 95 percent confidence intervals for these estimates. The dependent variable is the change in the specified log population head count between 2000 and the year indicated on the horizontal axis; the trade shock is the decadalized 2000–2012 change in CZ import exposure, as defined in equation (1) and instrumented by equation (3); and control variables are the same as in figure 5 plus CZ population growth over 1970–1990. Regressions are weighted by the CZ total population in 2000; standard errors are clustered by state.

the time interval lengthens. The other three panels of the figure make clear that these negative effects are driven by one age group, those 25–39 years old. For adults age 18–24 and 40–64, the impact of greater import competition on population head counts is negative but small and imprecisely estimated for all time differences. The finding of no net impact of trade exposure on CZ populations is consistent with Faber, Sarto, and Tabellini (2019), while the greater responsiveness of younger adults echoes Bound and Holzer (2000), who show that less experienced workers are relatively

15985-06a_Autor-5thPgs.indd 407 5/12/22 12:06 PM

more mobile in the face of adverse shocks. For the 2000–2019 time difference, the coefficient estimate of −5.89 (*t*-value = −2.52) indicates that when comparing CZs at the 25th and 75th percentiles of trade exposure, the more-exposed CZ would have a −3.89 (= −5.89 × [1.17 − .51]) percentage point larger decrease in head counts for this age group. For comparison, the 75th–25th percentile difference in population growth across CZs for individuals age 25–39 over 2000–2019 was 20.92 (= 16.08 + 4.84) percentage points. Cross-CZ spillovers transmitted via net migration appear to be modest and concentrated on a narrow age group.

The greater mobility of foreign-born workers may contribute to the labor supply responsiveness to the China trade shock seen in figure 7 (Cadena and Kovak 2016). To explore this channel, we use census and ACS data to evaluate the impact of trade exposure on population head counts by nativity. Online appendix figure A6 summarizes regressions for two time differences: 2000 (using data from the 2000 census) to 2010 (using data from the combined 2006–2010 ACS 1 percent samples), and 2000 to 2019 (using data from the combined 2015–2019 1 percent samples).[35] All other variables are the same as in the regressions in figure 7. The impact of trade exposure on the total working-age population is negative for both the 2000–2010 (β = −2.42, *t*-value = −1.92) and 2000–2019 (β = −3.47, *t*-value = −1.84) time differences, and marginally significant in each case. When we examine the impact on the native-born population, we find a smaller and much less precisely estimated effect: the impact coefficient for 2000–2019 is −0.75 (*t*-value = −0.79). This contrasts with impacts on foreign-born workers, for whom greater import competition significantly reduces population head counts over 2000–2019 (β = −6.79, *t*-value = −2.05). When looking at age subgroups within these populations, we see negative significant impacts of import competition on native-born workers age 25–39 (β = −5.03, *t*-value = −2.29), which aligns with the results in figure 7, and on foreign-born workers age 40–64 (β = −12.70, *t*-value = −2.08). Whereas native-born labor-supply responses are strongest for younger workers, for foreign-born workers they are strongest for older workers.

GRAVITY-BASED TRADE SHOCKS To evaluate how changes in labor-market outcomes in one region affect outcomes in other regions, Adão, Arkolakis, and Esposito (2019) build a general equilibrium trade model that captures such spillovers explicitly and generates reduced-form equilibrium

35. Because of small population counts of foreign-born workers in many CZs in any individual year, we use combined five-year ACS samples rather than the combined three-year samples used in online appendix figure A5.

conditions that have a shift-share structure. If national industries are subject to exogenous shocks, such as reform-induced growth in Chinese manufacturing, then employment and wages in regional economies will be affected through two channels. One is through changes in local industry revenue, which in the case of greater import competition will place downward pressures on local wages and employment, as captured by changes in import penetration in equation (2). Second, in the presence of cross-region spillovers, wages and employment in one region will also be affected by localized changes in import penetration in other regions, whether through cross-CZ trade or migration. For a given CZ, shocks to other CZs will matter more for employment outcomes the larger and the closer these other markets are, as dictated by the gravity structure of trade. Adão, Arkolakis, and Esposito (2019) quantify this cross-region spillover by adding to the specification in equation (2) the gravity-weighted changes in import competition in all other regions (i.e., the sum of the trade shock in each region weighted by the size of and the distance to that region).

We incorporate their approach by estimating an extended version of equation (2) that includes a gravity-based measure of trade exposure in other CZs. Online appendix A.5.3 provides details on the estimation and reports results. For all time differences, the impacts of trade shocks to a CZ's own industries are very similar in magnitude and significance to those in figure 4. The impacts of gravity-based trade shocks are small and imprecisely estimated, both for manufacturing employment and nonmanufacturing employment-population ratios. Evidence of gravity-based spillovers appears to be weak.

CROSS-INDUSTRY SPILLOVERS The literature also investigates spillovers in trade shocks between industries, which may operate both within and across regions. On within region transmission, Autor and others (2014) find that workers' earnings are adversely affected both by shocks to their own industry of employment and by shocks to the industries of other workers in the same CZ (as captured by the average trade shock for these workers' industries). Acemoglu and others (2016) and Pierce and Schott (2016) document spillovers that operate through input-output linkages in industry supply chains. Rising US furniture imports from China, for instance, may cause US factories in this downstream industry to reduce purchases of inputs from the upstream (i.e., supplier) sectors with which it is linked—for example, planed lumber, plywood, woodworking machinery, textiles, screws, and adhesives. Because buyers and suppliers often locate near one another, much of the impact of increased trade exposure in downstream industries may transmit to suppliers in the same regional market.

Using US input-output data to construct upstream (supplier) and downstream (customer) import exposure shocks for both manufacturing and nonmanufacturing industries, Acemoglu and others (2016) estimate negative employment effects in industries whose customer industries are directly trade-exposed. Conversely, they find little evidence for differential employment changes in industries whose suppliers are directly trade-exposed. Trade exposure thus appears to primarily propagate upstream in the supply chain—that is, from trade-exposed customers to their suppliers.

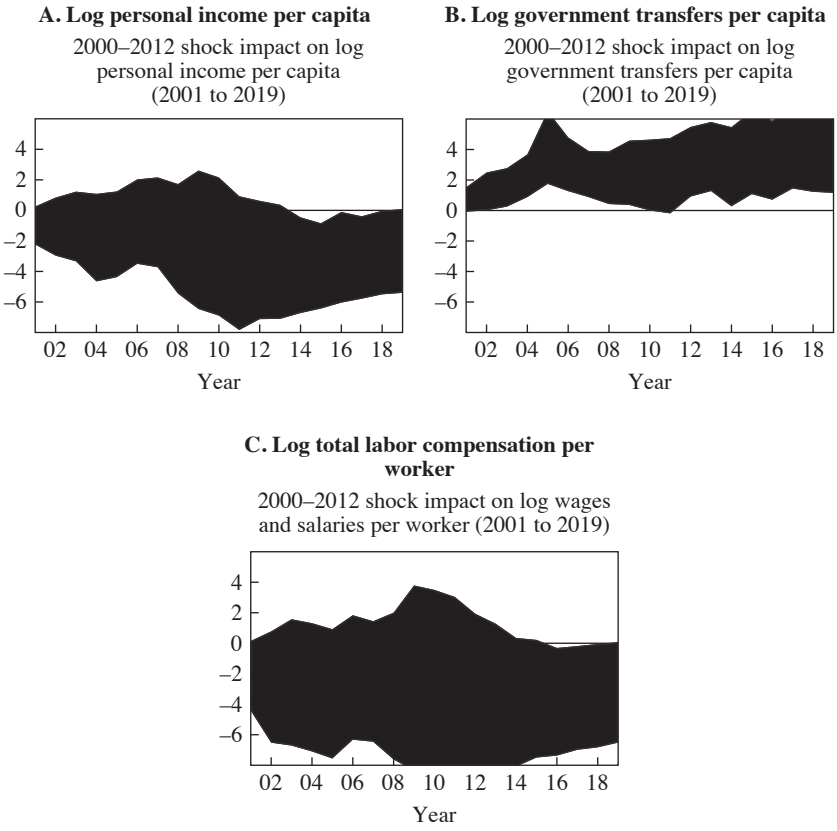
III.C. Personal Income, Labor Compensation, and Government Transfers

Adverse impacts of trade shocks on local labor demand are likely to reduce labor income, especially among low-wage workers. Chetverikov, Larsen, and Palmer (2016) and Autor, Dorn, and Hanson (2019) find that greater exposure to import competition from China caused larger relative reductions in earnings in the bottom four deciles of earnings distributions within CZs, where declines were larger for men than for women. Yet, impacts of trade exposure on labor earnings provide only a partial sense of the consequences for economic activity. Reductions in the employment-population ratio in a CZ may dampen demand for local goods and services and housing, possibly reducing revenues flowing to local business owners and landlords. Feler and Senses (2017) document negative impacts of the China trade shock on CZ housing values and property tax revenues, though they do not find evidence that CZs that are more trade-exposed saw larger declines in the number of business establishments. Declines in local income may in turn trigger greater uptake of government transfers, as more individuals qualify for means-tested government assistance or elect to retire and begin to receive Social Security income. Autor, Dorn, and Hanson (2013a) find that trade exposure caused an increase in government transfer receipts in adversely affected CZs, with most of the uptake accounted for by retirement, disability, and medical benefits. These prior analyses extend only through 2007, however, so they do not illuminate the long-run impact of the China trade shock on local area income.

To fill in this picture, we examine how trade exposure affects the components of personal income per capita. BEA estimates of local area personal income include labor compensation (wages, salaries, bonuses, employer contributions to health and pension plans), proprietor income (income of sole proprietorships, partnerships, tax-exempt cooperatives), financial returns (rent, interest, dividends, realized capital gains), government transfers (both cash and in-kind), and adjustments to capture income by place of residence.

15985-06a_Autor-5thPgs.indd 410 5/12/22 12:06 PM

Figure 8. Trade Shock Impact on Income, 2000–2019



Source: Authors' calculations.

Note: Panels report two-stage least squares coefficient estimates for β_{1h} in equation (2) and 95 percent confidence intervals for these estimates. The dependent variable is the change in the specified outcome between 2000 and the year indicated on the horizontal axis; the trade shock is the decadalized 2000–2012 change in CZ import exposure, as defined in equation (1) and instrumented by equation (3); and control variables are the same as in figure 5. Regressions are weighted by the CZ population in 2000; standard errors are clustered by state.

By including as many sources of income as possible, and by seeking to assign income according to the place of residence of recipients, BEA personal income per capita approximates aggregate local income per capita and is therefore suitable for evaluating the distributional consequences of trade exposure across CZs.

Estimation results for these outcomes appear in figure 8, where outcomes are in terms of log income and transfers relative to the total population of

a CZ, except for total labor compensation, which is relative to wage and salary employment.[36] The impact of exposure to import competition on personal income per capita in figure 8, panel A, is negative at all time horizons and precisely estimated for most time differences from 2000 to 2014 forward. Negative effects reach their peak for the 2000–2015 time difference, for which the impact coefficient is −3.64 (*t*-value = −2.55). The moderately attenuated coefficient estimate of −2.66 (*t*-value = −1.90) for the 2000–2019 time difference indicates that for CZs at the 25th and 75th percentiles of trade exposure, the more exposed CZ would have a 1.76 (= −2.66 × [1.17 − .51]) percentage point smaller increase in personal income per capita over 2000–2019, which compares to the 25th–75th percentile difference in the growth of personal income per capita over 2000–2019 of −12.2 (= 21.3 − 33.5) percentage points. Consistent with Autor, Dorn, and Hanson (2013a), we see in figure 8, panel B, that CZs that are more trade-exposed had larger increases in government transfer receipts per capita, which remain elevated through 2019. For the 2000–2019 time difference, the impact coefficient of 4.03 (*t*-value = 2.75) implies that at the 75th–25th percentile difference in trade exposure, the more exposed CZ would have a 1.81 (= 2.75 × [1.17 − .51]) percentage point larger increase in transfers per capita over the time period, relative to the 75th–25th percentile difference in the growth of transfers per capita of 10.90 (= 67.6 − 56.7) percentage points.

In online appendix A.5.4, we further explore the impact of trade shocks on government transfers by program type. Consistent with results in Autor, Dorn, and Hanson (2013a) for earlier time periods, adjustments in Social Security and Medicare benefits account for most of the responsiveness in government transfers induced by greater import competition, where the magnitude of these benefit gains expands as the time horizon lengthens. To receive these benefits, an individual must have left the labor force, either through retirement or by being declared medically unable to hold a job. The primary means through which government transfers replace labor income lost due to import competition is thus by accommodating an exit from paid work, which may help account for the long-run negative effects of trade exposure on employment-population ratios we see in figure 5. Despite trade-induced reductions in incomes, means-tested government programs

36. We normalize personal income by the total (rather than working-age) population, given that non-working-age individuals may earn income from nonlabor sources and receive government transfers of various kinds.

15985-06a_Autor-5thPgs.indd 412 5/12/22 12:06 PM

meant to provide income assistance to poor households are instead largely unresponsive to greater import competition.

III.D. Heterogeneity in Impacts across Regions

We have so far focused on the average response of CZs to import competition, which provides insight into the overall implications of the China trade shock. These average impacts mask heterogeneity in how CZs adjust to adverse shocks. Among the most trade-impacted CZs, there is wide variation in initial industrial specialization and labor-force composition (see online appendix table A4). Consider, for instance, North Hickory (2000 population 378,000), and Raleigh-Cary (2000 population 1,420,000), both of which are located in North Carolina. Over 2000–2012, the two CZs were each above the 95th percentile of trade exposure, with increases in import penetration of 4.4 percentage points for the industries in North Hickory and 3.4 percentage points for those in Raleigh-Cary. However, these CZs entered the 2000s with distinctly different economic structures. Hickory was a traditional factory town; in 2000, 43.0 percent of its working-age population had a manufacturing job and just 15.6 percent had a bachelor's degree. Raleigh-Cary, by contrast, was more educated and industrially diversified. In 2000, 34.2 percent of its working-age population had at least a college degree, and only 17.0 percent worked in manufacturing. These initial differences may have shaped how the two CZs responded to the large and sudden increase in import competition.

A substantial literature documents that US regions with fewer college-educated workers have grown less rapidly (Glaeser, Scheinkman, and Shleifer 1995; Behrens, Duranton, and Robert-Nicoud 2014; Diamond 2016) and have seen larger declines in employment-population rates (Austin, Glaeser, and Summers 2018). To explore the role that a dearth of human capital may play in the poor adjustment of local labor markets to the China shock, we separately examine CZs that initially had smaller versus larger supplies of college-educated workers.

Other work studies how greater specialization leaves regions more exposed to industry-specific shocks (Feyrer, Sacerdote, and Stern 2007; Michaels 2010). Many of the labor-intensive US industries exposed to trade with China were agglomerated in small and medium-sized cities in the Southeast and Midwest (Ellison, Glaeser, and Kerr 2010). Existing work documents that the impact of the China shock was greater in CZs that at the outset had lower employment rates (Austin, Glaeser, and Summers 2018) and were more specialized in mature industries positioned later in their

15985-06a_Autor-5thPgs.indd 413 5/12/22 12:06 PM

product cycles (Eriksson and others 2019).[37] We explore the role that reverse agglomeration contributes to poor adjustment to the China shock, and we separately analyze CZs that initially were less versus more industrially specialized.

Figure 9 presents estimates in which we divide the sample of CZs into groups based on whether the share of college-educated workers in their working-age populations was above or below the population-weighted national median in 2000. There are 336 CZs in the former group and 386 in the latter group. We consider four outcomes: the manufacturing employment-population ratio, the total wage and salary employment-population ratio, the log working-age population, and log personal income per capita. To keep the time horizon constant across outcomes, we evaluate time differences from 2001–2002 to 2001–2019. To control the false discovery rate when evaluating differences in coefficient estimates across sample splits, we compute and display minimal Benjamini-Hochberg q-values based on the number of hypotheses being tested.[38]

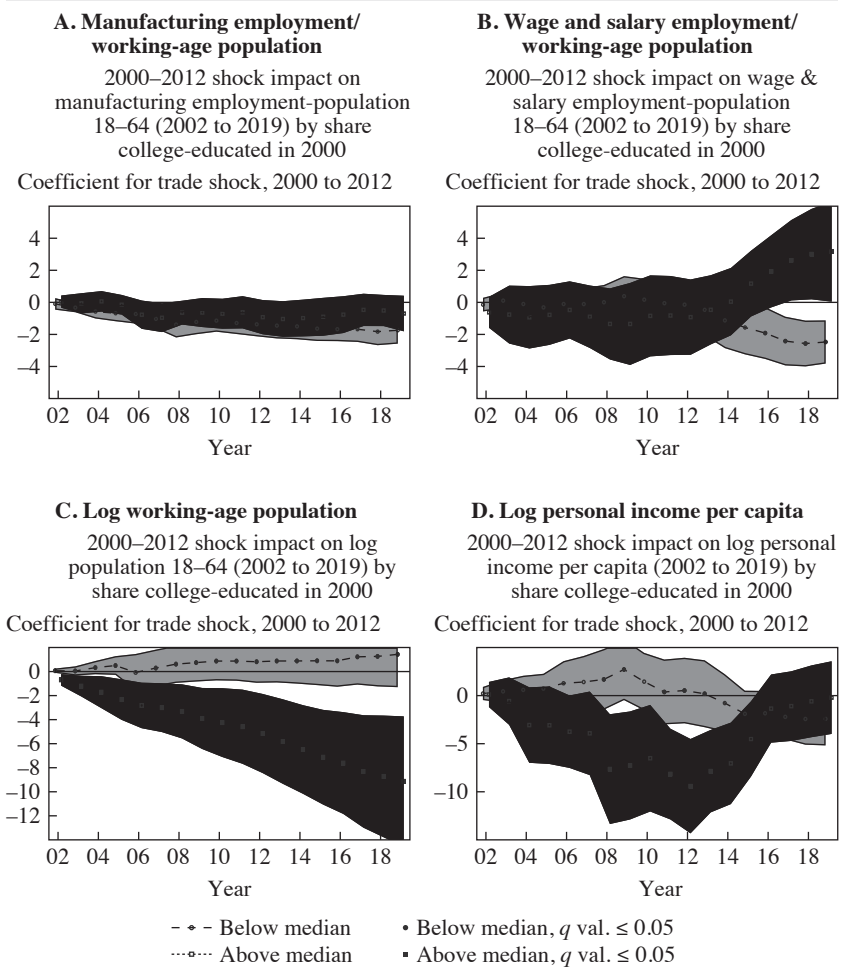
Although CZs with higher-educated and lower-educated workers both experienced declines in manufacturing employment, the negative impacts of trade exposure on overall employment in figure 5 are concentrated in CZs with relatively few college-educated workers, as shown in figure 9, panels A and B. These results are consistent with those in Bloom and others (2019), who study the 1992–2012 time period. For the 2001–2019 horizon, the first two figures show that in CZs with less-educated workers, a 1 percentage point increase in import penetration over 2000–2012 predicts a 1.74 (t-value = –4.07) percentage point decrease in the manufacturing employment-population ratio and a 2.47 (t-value = –3.59) percentage point decrease in the total wage and salary employment-population ratio. Across all CZs (see figure 5), trade-induced changes in manufacturing and total employment-population ratios are very similar at long horizons, indicating

37. In Brazil, Dix-Carneiro and Kovak (2017) document that regional manufacturing continued to decline well after trade reform had permitted greater import competition, in a manner consistent with agglomeration economies.

38. The number of hypotheses (144) is the product of the four outcomes, eighteen years in the sample, and two sample splits. In the standard Benjamini-Hochberg procedure, hypotheses are ranked according to their unadjusted p-value. For a fixed significance level q, the researcher rejects all hypotheses that satisfy $p < qr / M$, where p is the p-value, r is the rank of the p-value, and M is the number of hypotheses being tested. Following the step-up method in Anderson (2008), we find the smallest q at which each null hypothesis can be rejected. In figures 8 and 9, we show the coefficients whose differences across sample splits have a minimal q-value of less than 0.05 with solid markers and use hollow markers everywhere else.

15985-06a_Autor-5thPgs.indd 414 5/12/22 12:06 PM

Figure 9. Heterogeneity in Trade Shock Impacts: Initial College-Educated Population



Source: Authors' calculations.

Note: Panels report two-stage least squares coefficient estimates for β_{1h} in equation (2) and 95 percent confidence intervals for these estimates. Coefficient estimates whose differences have a minimal Benjamini-Hochberg q-value of less than or equal to 0.05 are shown with solid markers (with hollow markers for other estimates). Estimates are reported for two samples: the 386 CZs whose share of college-educated workers in the working-age population was below the population-weighted national median in 2000, and the complementary set of 336 CZs. The dependent variable is the change in the indicated measure between 2001 and the year on the horizontal axis; the trade shock is the decadalized 2000–2012 change in CZ import exposure as defined in equation (1) and instrumented by equation (3); control variables are the same as in regressions reported in figure 4. Regressions in panels A–C are weighted by the CZ working-age population in 2000; regressions in panel D are weighted by the CZ total population in 2000. Standard errors are clustered by state.

15985-06a_Autor-5thPgs.indd 415 5/12/22 12:06 PM

that employment impacts on nonmanufacturing employment are null. For CZs with relatively few college-educated workers, the substantially larger impact on total employment than on manufacturing employment reveals a negative impact of trade exposure on nonmanufacturing employment, as shown in online appendix figure A9. Despite manufacturing job losses being compounded by nonmanufacturing losses in these CZs, there is no effect of trade exposure on the size of the log working-age population (figure 9, panel C). It thus appears that more trade-impacted CZs with fewer college-educated workers did not experience differential out-migration, though they did experience larger declines in personal income per capita, as shown in figure 9, panel D. Since economically motivated migration would tend to mitigate local employment-reducing effects of trade exposure, one interpretation of these results is that the lack of migration is both symptom and cause of the slow adjustment process.

In CZs with more-educated working-age populations, the pattern of adjustment is qualitatively different. Impacts of trade exposure on manufacturing employment (figure 9, panel A) are negative but somewhat smaller and less precisely estimated. Impacts on the total wage and salary employment-population ratio are small and imprecise for short time differences and then become large, positive, and marginally significant for long time differences. For the 2001–2016 time difference and beyond, we easily reject that impact coefficients for wage and salary employment-population are the same in CZs with more-educated versus less-educated workers. This trade-induced increase in the total employment-population ratio must imply a corresponding positive impact of trade exposure on the nonmanufacturing-population ratio, as seen in online appendix figure A9.

Just as critical, CZs with more-educated workers adjusted to adverse trade shocks in part through net out-migration. For all time differences except 2001–2006, we reject that impact coefficients on the working-age population are equal for CZs with more-educated versus less-educated workers. At the 2001–2019 time difference in CZs with more-educated workers, a 1 percentage point increase in import penetration is predicted to cause a 9.13 (t-value = –3.31) percentage point decrease in the working-age population, or an annual population decline of approximately one-half a percentage point. The impact coefficient on log nonmanufacturing employment at the 2001–2019 time difference of 6.04 (t-value = 2.37, online appendix figure A9, panel b) implies that more of the increase in the nonmanufacturing employment population ratio occurred through the out-migration of labor rather than through greater job growth. Like CZs with less-educated workers, those with more-educated workers also see negative

15985-06a_Autor-5thPgs.indd 416 5/12/22 12:06 PM

impacts of trade exposure on personal income per capita (figure 9, panel D). Distinct from CZs with less-educated workers, these impacts reach their maximum negative value for the 2001–2012 time difference and then diminish over time, becoming close to zero for the 2001–2016 horizon and beyond.[39]

Next, we consider a second dimension of heterogeneity in adjustment to the China trade shock. In figure 10, we split the sample of CZs by their industrial specialization in 2000, which we measure using a Hirschman Herfindahl Index (HHI) equal to the sum of the squared shares of CZ employment in each industry.[40] There are 103 CZs with HHIs above the population-weighted median and 619 CZs with HHIs below it. The uneven split in CZs across the two groups reveals, unsurprisingly, that large CZs are less specialized by industry than smaller and medium-sized CZs. In figure 10, panels A and B, it appears that the negative impacts of trade exposure on manufacturing and wage and salary employment across all CZs are driven predominantly by smaller, more industrially specialized CZs. For these CZs, there are negative and significant impacts of greater import competition on manufacturing employment and total wage and salary employment-population ratios. For less specialized CZs, impacts on manufacturing employment are small, negative, and imprecise, and impacts on total employment are null. In the more-specialized CZs but not in the less specialized ones (figure 10, panel C), trade exposure induces larger decreases in the working-age population. Despite this, employment-population ratios decline by relatively more in more trade-exposed and more specialized CZs. Our conclusions about differences in adjustment between more and less specialized CZs are tentative, however, given that for most time horizons and outcomes we cannot reject that impact coefficients for the two groups of CZs are the same.

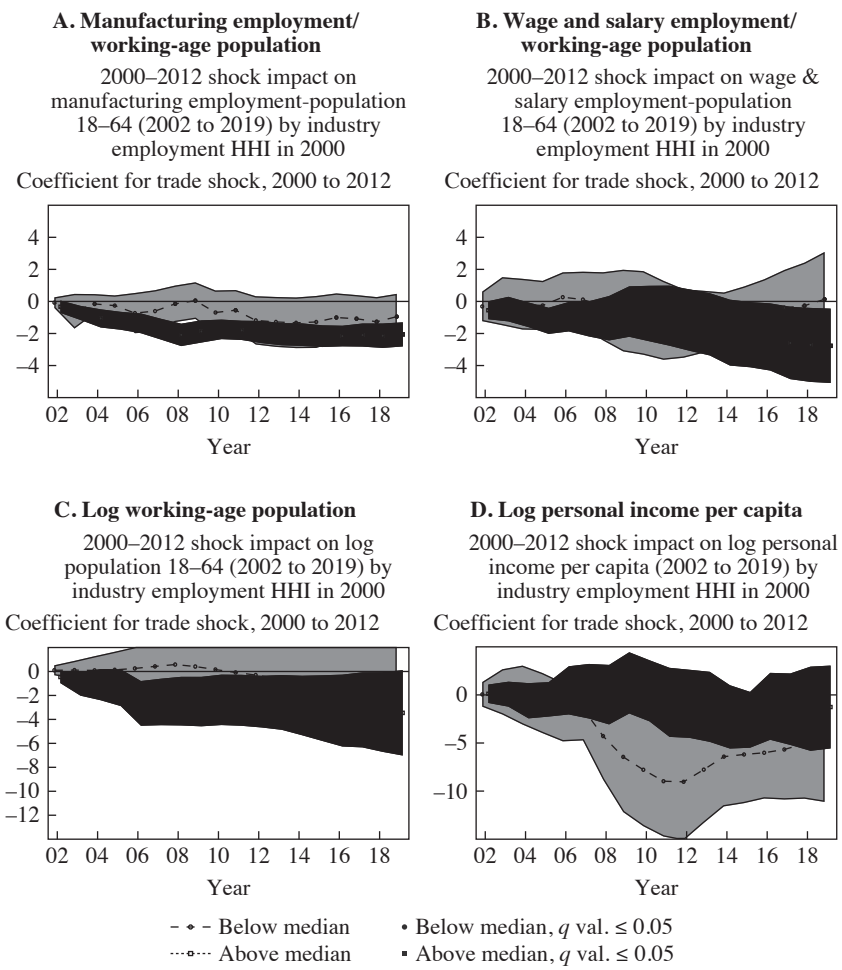
III.E. Putting the Pieces Together

The first wave of China shock research found that greater import competition caused localized job loss in manufacturing, declines in earnings for low-wage workers, and greater economic distress across a wide range of outcomes. The primary mechanism of adjustment to trade exposure was exit from work, rather than increased employment in nonmanufacturing or

39. Online appendix figure A9, panels c and d, show that in both CZs with more-educated and less-educated workers trade exposure induced an increase in government transfers per capita, with the majority of this increase accounted for by payments related to Social Security and Medicare.

40. We construct these HHIs using ACS data for 2000 and the 1990 census industry classification code.

Figure 10. Heterogeneity in Trade Shock Impacts: Initial Industry Specialization



Source: Authors' calculations.

Note: Panels report two-stage least squares coefficient estimates for β_{1h} in equation (2) and 95 percent confidence intervals for these estimates. Coefficient estimates whose differences have a minimal Benjamini-Hochberg *q*-value of less than or equal to 0.05 are shown with solid markers (with hollow markers for other estimates). Estimates are for two samples: the 619 CZs whose Hirschman Herfindahl Index of industry specialization was above the population-weighted national median in 2000, and the complementary set of 103 CZs. The dependent variable is the change in the indicated measure between 2001 and the year on the horizontal axis; the trade shock is the decadalized 2000–2012 change in CZ import exposure as defined in equation (1) and instrumented by equation (3); control variables are the same as in regressions reported in figure 4. Regressions in panels A–C are weighted by the CZ working-age population in 2000; regressions in panel D are weighted by the CZ total population in 2000. Standard errors are clustered by state.

15985-06a_Autor-5thPgs.indd 418 5/12/22 12:06 PM

migration to other regions. The impacts of rising import competition, originally documented for 1991 through 2007, are manifest out to 2019, nearly two decades after China's accession to the WTO in 2001 and nine years after the plateau of China's export surge. CZs that are more trade-exposed suffered durable increases in joblessness and decreases in personal income per capita that are not close to being offset by government transfers. The resulting economic distress appears to be most acute in local labor markets that lacked abundant supplies of college-educated workers—consistent with the dearth of human capital hypothesis—and that were narrowly specialized in labor-intensive manufacturing—consistent with the reverse agglomeration hypothesis. For CZs such as North Hickory, North Carolina, the consequences of the China trade shock have been profound and long-lasting.

There are other plausible explanations for the lack of an out-migration response to the China trade shock. These include downward pressure on housing values in contracting CZs (Feler and Senses 2017), which for renters may have diminished the pressure to leave and for homeowners may have complicated selling their homes (Glaeser and Gyourko 2005; Notowidigdo 2020). Lower housing values may have further blocked local recovery by slowing the formation of new businesses, which are often financed using home equity (Davis and Haltiwanger 2019).[41] It does not appear, however, that regional variation in labor market regulations account for differential adjustment to the China shock. The impacts of trade exposure were no less acute in CZs located in right-to-work states or states with lower minimum wages (Chan 2019).

There are alternative characterizations of the China trade shock that deserve mention. One is that, amid the regular ebb and flow of the US labor market, the China shock was no more consequential than a high tide. If every year millions of jobs are created and millions of jobs are destroyed, how could job loss in a collection of factory towns be important? A second is that manufacturing job loss would have happened anyway.[42] That is, import competition from low-wage countries other than China (Hanson 2020),

41. Housing market regulations in large US cities could also have impeded adjustment in trade-exposed regions. Inelastic housing supply in major cities, due in part to housing regulations (Glaeser, Gyourko, and Saks 2005), may hinder low-wage workers in the heartland from moving to expensive coastal cities (Hsieh and Moretti 2019).

42. A related argument, that it was technological change and not import competition that caused manufacturing job loss in trade-exposed CZs, has two weaknesses. One is that the cross-CZ correlation between exposure to automation and exposure to import competition is very low (Autor, Dorn, and Hanson 2013b); a second is that there was no obvious abrupt acceleration in technological change after 2001 (indeed, productivity growth slowed after 2004).

15985-06a_Autor-5thPgs.indd 419 5/12/22 12:06 PM

rising capital intensity in manufacturing (Fort, Pierce, and Schott 2018), and the availability of industrial robots (Acemoglu and Restrepo 2020) would ultimately have closed factories and displaced workers, irrespective of China's rise. Both lines of thought overlook how concentrated the China trade shock was in place and time. The concentration in place was due to the relocation of manufacturing to smaller cities and towns over the twentieth century, made possible by the maturation of industrial production (Eriksson and others 2019) and improved transportation (Kim 1995; Michaels 2008). The concentration in time was due to the speed of China's rise and the immensity of its economy (Naughton 2006). The scarring effects of job loss in trade-exposed CZs were made more acute by this job loss occurring during trade-induced local economic recessions (Jacobson, LaLonde, and Sullivan 1993; Davis and von Wachter 2011; Huckfeldt 2021). Workers were not just losing their jobs, they were experiencing the dislocating effects of work disappearing in their communities (Wilson 1996; Autor, Dorn, and Hanson 2019; Charles, Hurst, and Schwartz 2019).

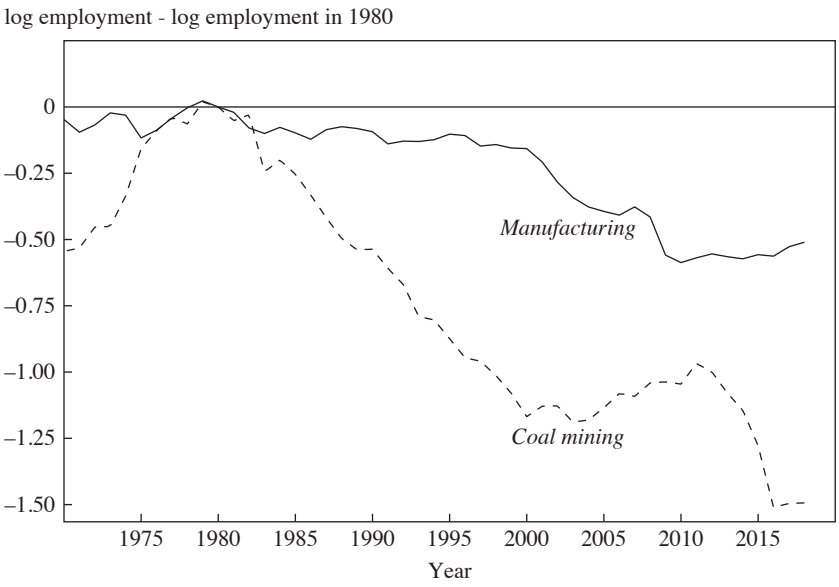
IV. Was the China Shock Unique?

A question arises as to whether the China trade shock was a singular event with limited relevance for other shocks or whether it holds general lessons about concentrated job loss. The regional consequences of trade-induced employment contractions in the United States mirror those of the broader decline of manufacturing. Charles, Hurst, and Schwartz (2019) document that US CZs suffering greater job loss in manufacturing after 2000—whether because of import competition, automation, or other forces—had larger declines in employment rates and wages out to 2017. We do not know, however, whether this pattern of adjustment is closely comparable to shocks that are not specific to manufacturing or that occurred at other moments in time.

To provide a benchmark against which to compare the China trade shock, we examine outcomes associated with two other major events whose impacts were also highly localized. One is the decline of the coal industry, which after 1980 suffered two decades of contraction due to changes in production technology and energy demand (Black, Daniel, and Sanders 2002; Black, McKinnish, and Sanders 2005).[43] A second is the Great Recession, in which

43. On regional boom and bust cycles in energy production, see Michaels (2010), Jacobsen and Parker (2014), Allcott and Keniston (2017), Feyrer, Mansur, and Sacerdote (2017), Bartik and others (2019), Clay and Portnykh (2018), and Kearney and Wilson (2018).

Figure 11. Employment Decline in Coal Production versus Manufacturing



Sources: BEA Industry Economic Accounts and authors' calculations.

a housing collapse and subsequent credit freeze differentially affected US regions (Mian and Sufi 2014; Charles, Hurst, and Notowidigdo 2018; Beraja, Hurst, and Ospina 2019; Yagan 2019). The first episode lets us evaluate long-run adjustment to a sectoral shock during an earlier period, while the second involves adjustment to an unusually deep cyclical shock.

IV.A. Labor Market Adjustment to the Decline of Coal

The secular decline of employment in the coal industry is a case of a spatially concentrated shock that precedes China's rise. Figure 11 plots employment in US coal mining and, for reference, employment in US manufacturing, each expressed as the log for a given year relative to the log value in 1980. There was a boom in coal production during the 1970s due to a spike in energy prices caused by the decade's two major oil price shocks. Employment in coal rose from 140,600 workers in 1969 to 263,600 workers in 1979. After 1980, coal prices began an extended decline. Employment in coal followed suit, falling to 151,200 workers in 1990 and to 80,400 workers in 2000. After rebounding modestly in the 2000s, coal employment plummeted after 2011 (Black, McKinnish, and Sanders 2005). Given these

patterns, we specify the coal shock as the change in coal employment over 1980–2000. Our analysis builds on Black, McKinnish, and Sanders (2005), who find that during the 1983–1993 coal bust, counties more specialized in coal (as indicated by initial coal reserves) in Kentucky, Ohio, Pennsylvania, and West Virginia had larger decreases in employment rates, total earnings, and earnings per worker. We expand their analysis forward in time from 1993 to 2019 and include all US CZs.[44]

To evaluate local labor market adjustment to the decline of coal, we adapt the specification in equation (2) to the following:

$$\Delta Y_{it+h} = \alpha_t + \beta_{1h}\Delta SS_i^{coal} + \mathrm{X}'_{it}\beta_2 + \varepsilon_{it+h}, \tag{4}$$

where ΔY_{it+h} is the change in an outcome for CZ i between year $t + h$ for $h = -5, \ldots, 39$ and the base year $t = 1980$; ΔSS_i^{coal} is a shift-share variable that projects the coal shock onto CZ i; and X_{it} is a vector of controls. We specify the decadalized shift-share coal shock as,

$$\Delta SS_i^{coal} = -\left(\frac{100}{20}\right) \times \frac{L_{ic1980}}{L_{i1980}}\left[\ln L^{-i}_{coal,2000} - \ln L^{-i}_{coal,1980}\right], \tag{5}$$

where L_{ic1980}/L_{i1980} is the share of coal production in the employment of CZ i in 1980, and the term in brackets is the log change in national employment in coal production over 1980–2000, outside of the state in which CZ i is located. To facilitate interpretation, in the regression analysis we multiply the shock in equation (5) by −1, such that a higher value of the shift-share variable indicates a larger negative change. The population-weighted mean of this projected employment change is 0.22 ($\sigma = 1.25$) percentage points.[45] Control variables in the regression include a dummy for the CZ having positive coal employment in 1980, time trends for census region divisions, and, parallel to above, values in 1980 for the share of CZ employment in manufacturing, the share of women in CZ employment, the share of college-educated workers in the CZ population, and the share of foreign-born workers in the CZ population. Because of the spatial concentration of coal deposits, the shock hit a relatively small number of CZs: just 258 CZs

44. Because CZs are aggregates of counties, our analysis may dampen the county-level effects in Black, McKinnish, and Sanders (2005).

45. Given our shock definition, this indicates that the average change in coal employment across CZs was negative. The correlation between the coal shift-share variable in equation (5) and the China shock shift-share variable in equation (2) is –0.08.

15985-06a_Autor-5thPgs.indd 422 5/12/22 12:06 PM

(35.7 percent) had positive coal employment in 1980, and the population-weighted median value of the shock is only 0.001 percentage points. By contrast, the shocks at the 90th and 95th percentiles were 0.19 and 0.83 percentage points, respectively.

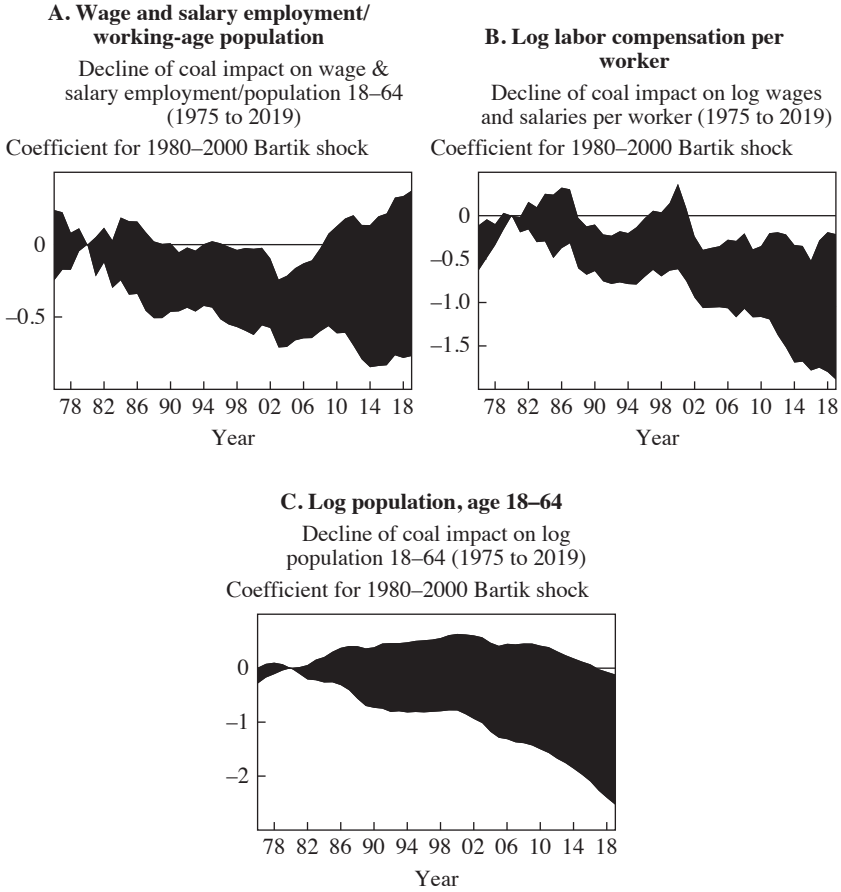
Estimation results appear in figure 12. The impact coefficients for the ratio of wage and salary employment to population (figure 12, panel A) become increasingly negative during the 1980–2000 shock. They are largest for the 1980–2003 horizon (−0.48, *t*-value = −3.98) and remain negative for all horizons out to 1980–2019, though the estimates for the longest time windows lack statistical precision. For the 1980–2006 difference—on the eve of the Great Recession and twenty-six years after coal's decline had begun—the impact coefficient of −0.39 (*t*-value = −2.94) indicates that a CZ subject to a 1980–2000 coal shock at the 95th percentile of intensity would have experienced a 0.32 (= 0.39 × [0.83 − .001]) percentage point larger decline in its wage and salary employment-population ratio than a CZ at the median of shock intensity. For context, over 1980–2006 the 5th–50th percentile difference in the change in employment-population was −9.4 percentage points.

In figure 12, panel B, we see that CZs more exposed to the decline of coal also had larger reductions in total labor compensation per worker, which persist well beyond the 1983–1993 period studied in Black, McKinnish, and Sanders (2005) and are precisely estimated at long horizons. Over 1980–2019 ($\beta = -1.05$, *t*-value = −2.46), a CZ subject to a coal shock at the 95th percentile of intensity would have seen a 0.88 (= −1.05 × [0.83 − .001]) log point greater decline in average labor income.

Given these lasting adverse labor market impacts of the coal shock, one might expect that heavily exposed CZs saw substantial population declines. However, figure 12, panel C, shows that the impact of exposure to the coal shock on population head counts is small and imprecisely estimated for the first two decades after the coal contraction began. Larger negative population responses appear only well after 2000 and are statistically significant only for time periods beyond 1980–2016. For the 1980–2019 difference, the impact coefficient of −1.33 (*t*-value = −2.15) indicates that a CZ subject to a coal shock at the 95th percentile of intensity would have had a 1.22 (= −2.23 × [0.83 − .001]) log point larger decline in its working-age population when compared to a CZ at the median of shock intensity. These impacts are small and do not appear until twenty-five years after the shock arrived.

Qualitatively, figure 12 indicates that the reaction of employment, earnings, and population to the coal shock was similar to the response to the

Figure 12. Local Labor Adjustment to the Decline of Coal



Source: Authors' calculations.

Note: Panels report two-stage least squares coefficient estimates for β_{1h} in equation (4) and 95 percent confidence intervals for these estimates. The dependent variable is the change in the indicated measure between 1980 and the year on the horizontal axis; the coal shock is defined in equation (5); control variables in the regression include a dummy for the CZ having positive coal employment in 1980, time trends for census region divisions, and values in 1980 for the share of CZ employment in manufacturing, the share of women in CZ employment, the share of the college educated in the CZ population, and the share of the foreign born in the CZ population. Regressions are weighted by the CZ working-age population in 1980; standard errors are clustered by state.

15985-06a_Autor-5thPgs.indd 424 5/12/22 12:06 PM

China import shock (see figures 5, 7, and 8). In both cases, CZs with greater exposure to the adverse sectoral shocks suffered declines in employment and earnings that persisted well beyond the period reaching peak intensity. These local labor markets eventually saw a decline in population but that decline was neither immediate nor large. Despite these qualitative similarities between the coal shock and the China trade shock, there is a substantial quantitative difference between the two. Whereas the estimates in figure 5 suggest that the China shock induced a decline in the wage and salary employment-population ratio of −1.55 (= −1.74 × 0.89) percentage points between 2001 and 2019, the corresponding employment decline for the coal shock was only −0.04 (= −0.198 × 0.22) percentage points per decade from 1980 to 2019.[46] The much smaller average impact of the coal shock may be unsurprising given that only a subset of CZs produce coal. Nor did the coal shock generate the extremes of variation in labor market outcomes across CZs that we observe for the China trade shock. Although wage and salary income per worker in a CZ at the 95th percentile of exposure to the coal shock declined by 0.88 log points more than in a CZ at the 5th percentile over the 1980–2019 period, the differential impact of the China shock on wages from 2000 to 2019 was three times as large when comparing CZs at the 75th and 25th (instead of 95th and 5th) percentile of exposure.

IV.B. Labor Market Adjustment to the Great Recession

As a second point of comparison, we evaluate local labor market adjustment to the Great Recession. We adapt the specification in equation (4) where ΔY_{it+h} becomes the change in an outcome for CZ i between year $t + h$ for $h = -5, \ldots, 13$, and base year $t = 2006$; ΔSS_i^{gr} is a shift-share variable that projects the Great Recession onto CZ i; and X_{it} is a vector of controls. The National Bureau of Economic Research (NBER) dates the Great Recession as beginning in December 2007 and ending in June 2009. Because US housing markets, which played a major role in the downturn, began contracting in 2006 (Charles, Hurst, and Notowidigdo 2016), we specify the shock as commencing in that year. We estimate equation (4) for twenty-one separate time periods: five periods before the shock begins (2001–2006 to 2005–2006) to check for pre-trends in outcomes; three shock periods

46. We obtain these estimated impacts by multiplying the impact coefficient for the wage and salary employment/working-age population ratio (−1.74 for the China shock, −0.19 for the coal shock) that corresponds to the longest time difference (2001–2019 for the China shock, 1980–2019 for the coal shock) times the decadalized mean of the respective shock (0.89 for the China shock, 0.22 for the coal shock).

15985-06a_Autor-5thPgs.indd 425 5/12/22 12:06 PM

(2006–2007 to 2006–2009) to examine shock impacts as the shock intensifies; and thirteen periods after the recession had ended (2006–2007 to 2006–2019) to evaluate longer-run adjustment.

Similar to equation (5), we specify the annualized Great Recession shock in shift-share form as,

$$\Delta SS_i^{gr} = -\left(\frac{100}{3}\right) \times \sum_j \frac{L_{ij2000}}{L_{i2000}} \left[\ln L^{-i}_{j,2009} - \ln L^{-i}_{j,2006}\right], \tag{6}$$

where L_{ij2000}/L_{i2000} is the share of industry j in employment of CZ in the year 2000, and the term in brackets is the log change in national employment in industry j over 2006–2009, outside of the state in which CZ i is located. To help interpret the results, we again multiply the shock in equation (6) by −1, such that a higher value indicates a larger negative projected change in employment. The population-weighted mean of the shift-share variable is 2.19 ($\sigma = 0.73$) percentage points—on average, CZs faced strongly contractionary forces during the recession period. The control variables included in the regressions are the same as those in the China shock regressions in equation (2).[47]

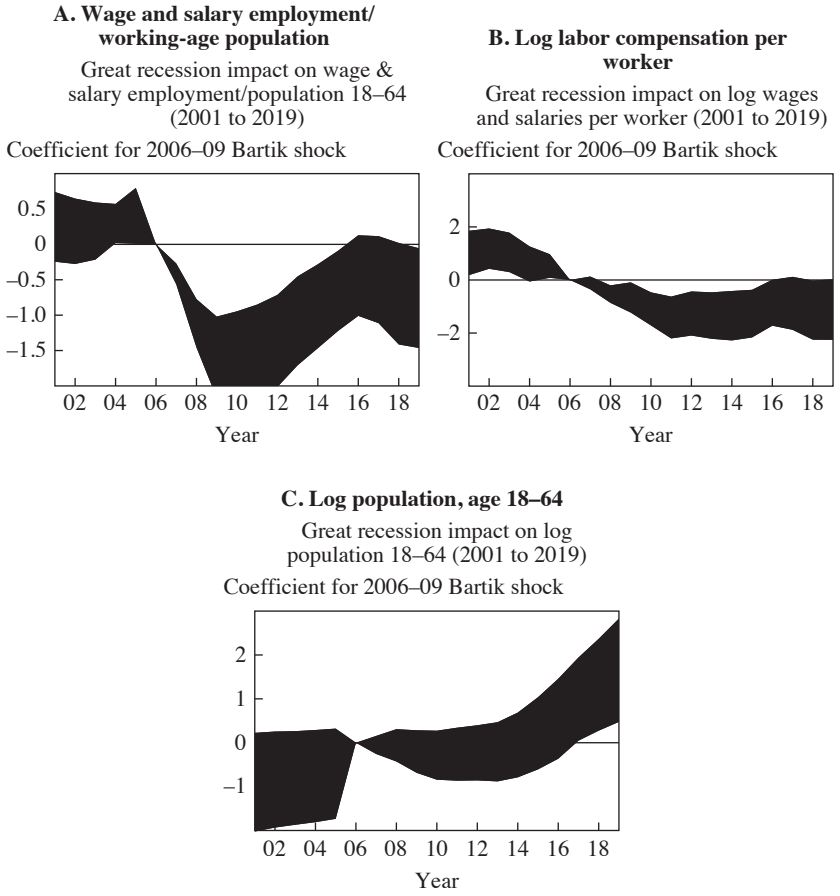
Yagan (2019) documents that the recovery in unemployment rates after the Great Recession was not matched by a recovery in employment rates. In CZs more adversely affected by the downturn, employment rates were suppressed out to 2015. Charles, Hurst, and Notowidigdo (2016) document further that declines in employment rates were concentrated among less-educated workers. We extend Yagan's analysis out to 2019, using CZ-level REIS data, as compared to Yagan's use of federal income tax records.

Estimation results appear in figure 13. CZs more exposed to contracting industries during the Great Recession saw larger declines in their employment rates (figure 13, panel A). Impact coefficients for the wage and salary employment-population ratio become rapidly more negative as the recession intensified, reaching their peak value for the 2006–2009 difference ($\beta = -1.60$, t-value = −5.42). As the post-2009 expansion began, CZs began to recover from earlier declines in employment; the impact coefficient for the 2006–2016 period is less than one-third the size of the 2006–2009 effect.

47. The correlation between the Great Recession shift-share variable in equation (6) (over 2006–2009) and the China shock shift-share variable in equation (2) (over 2000–2012) is 0.48. In unreported robustness exercises, we replicate the results in figure 13 in which we include as a control variable the instrument for the China trade shock, as defined in equation (3). Coefficient estimates and patterns of significance for the two sets of results are very similar.

15985-06a_Autor-5thPgs.indd 426 5/12/22 12:06 PM

Figure 13. Local Labor Market Adjustment to the Great Recession



Source: Authors' calculations.

Note: Panels report two-stage least squares coefficient estimates for β_{1h} in equation (4) and 95 percent confidence intervals for these estimates. The dependent variable is the change in the indicated measure between 2006 and the year on the horizontal axis; the Great Recession shock is defined in equation (6); control variables include initial period CZ employment composition (shares of employment in manufacturing, routine-task-intensive occupations, and offshorable occupations, as well as employment share among women), initial period CZ demographic conditions (shares of the college educated, the foreign born, non-white individuals, and those age 0–17, 18–39, and 40–64 in the population), and census region dummies. Regressions are weighted by the CZ working-age population in 2001; standard errors are clustered by state.

This recovery appears to stall in 2016, however. For the 2006–2019 long difference, CZs that were more exposed to the Great Recession still register net declines in their employment-population ratios. The impact coefficient of –0.76 (*t*-value = –2.11) at that horizon indicates that a CZ subject to the mean Great Recession shock over 2006–2009 would have experienced a 1.66 (= 2.19 × 0.76) percentage point decline in its wage and salary employment-population ratio, relative to the mean change for the 2006–2019 period of +2.20 percentage points ($\sigma = 3.89$).

In figure 13, panel B, we see that CZs more exposed to the Great Recession had larger declines in wage and salary income. The impact reaches its maximum for the 2006–2011 time difference ($\beta = -1.41$, *t*-value = –3.52), then attenuates modestly through 2016, after which recovery stops. Over the 2006–2019 long difference, a CZ subject to the mean Great Recession shock would have experienced a 2.45 (= –2.19 × 1.12) percentage point decline in compensation per worker.

The enduring declines in employment ratio and labor income per capita over the period 2006–2019 could indicate persistent scarring effects of the Great Recession. However, figure 13, panels A and B, also show that CZs with greater exposure to the Great Recession had differentially faster growth in employment and labor income leading up to the recession. To the extent that this faster growth was driven by a temporary boom in construction that expanded employment and earnings to unsustainable levels (Charles, Hurst, and Notowidigdo 2016), one might expect that these variables would not fully recover following the recession. Indeed, the estimates in figure 13, panel C, for population belie the expectation of a lasting downturn in the CZs with greater exposure to the Great Recession. These CZs experienced only modest and imprecisely estimated population declines during the recession and, from 2006 to 2016 onward, had faster relative population growth. This rapid growth raises the denominators for both the employment-population ratio and average wage and salary income, thus helping to explain why the recovery of those outcomes stalls in exposed CZs after 2015. These CZs were rapidly gaining population, which is likely a positive economic development.

In summary, we observe that Chinese import competition, the fall in demand for coal, and the Great Recession each reduced employment and labor incomes as these shocks unfolded. CZs with greater shock exposure did not experience large declines in population in any of these cases, suggesting that migration did not disperse the local impacts of these shocks. CZs more exposed to the Great Recession eventually experienced a substantial recovery of their employment rates and faster population growth. Those with

15985-06a_Autor-5thPgs.indd 428 5/12/22 12:06 PM

greater exposure to the China or coal shocks instead endured persistently depressed employment rates and labor income levels, combined with a slow decline in population. The comparison with the coal shock in particular indicates that the China shock's long-lasting impact on CZ labor market conditions and the sluggish population response to depressed local labor market conditions was not without precedent. However, the large magnitude of the China shock, which had a sizable impact on many local labor markets in the United States, sets it apart from shocks such as the decline of the coal sector, the impact of which was more limited in scope. As the United States prepares for potentially more job loss due to ongoing energy transformation and expected changes in oil and gas production, the failure of local labor markets to adjust successfully to the coal and China trade shocks reminds us that the adjustment process is typically slow and sclerotic, unlike the textbook model of frictionless labor market adjustment.

V. The Distributional Consequences of the China Trade Shock

Local labor markets more exposed to import competition from China suffered larger declines in manufacturing jobs, employment-population ratios, and personal income per capita. These effects persist for nearly two decades beyond the intensification of the trade shock after 2001, and almost a decade beyond the shock reaching peak intensity. The absence of a substantial out-migration response to the shock indicates that resident populations in exposed CZs bore the brunt of local job loss and its repercussions. While it is uncontroversial to infer that trade-exposed CZs suffered losses in average economic well-being *relative to* less trade-exposed CZs, these results do not reveal which if any CZs experienced absolute declines in average welfare or by how much. Computing absolute effects requires combining reduced-form empirical analysis, which estimates relative effects, with a general equilibrium model, which quantifies aggregate gains.

V.A. Quantitative General Equilibrium Analysis of the China Trade Shock

For individuals living in a given US CZ, trade with China affects well-being by changing earnings and consumption possibilities. If economic reform in China raises its productivity in labor-intensive manufacturing, economies with a comparative advantage in these industries would see demand for their factor services contract, thus diminishing their consumption possibilities, while economies whose comparative advantage lies elsewhere would see

their factor demand and consumption possibilities expand (Arkolakis, Costinot, and Rodríguez-Clare 2012). If labor is fully mobile across regions and sectors within a country, as has traditionally been assumed in modeling trade shocks, then the change in welfare would be common across regions. The evidence above contradicts such baseline assumptions; to a startling degree, trade shocks appear to have an enduring impact on the locations in which their immediate impact is felt. To interpret cross-region differences in welfare impacts emanating from trade shocks, theoretical models must incorporate frictions—normally labor market frictions—that produce the concentrated geographic impacts that we observe in the data.

As a benchmark for how labor immobility affects the welfare impact of the China trade shock, consider the analysis in Galle, Rodríguez-Clare, and Yi (2017).[48] They combine a model of heterogeneous firms from Eaton and Kortum (2002) with a model of heterogeneous workers from Roy (1951).[49] Workers are immobile across regions and partially mobile across sectors. They specify the China trade shock as innovations to China's industrial productivity, which they calibrate for the period 2000–2007 by backing out the implied productivity growth needed to generate the observed change in US manufacturing imports from China that is predicted by the first-stage regression of Autor, Dorn, and Hanson (2013a). Using these innovations, plus assumptions on other parameter values, they calculate that the average change in real income across CZs from trade with China is 0.22 percent, with a (unweighted) standard deviation across CZs of 0.31 percentage points.

The analysis in Caliendo, Dvorkin, and Parro (2019), also based on Eaton and Kortum (2002), allows labor to be mobile geographically in the longer run. Dynamic adjustment in labor allocations across US states and sectors arises from workers facing a fixed cost in moving from one region-industry to another and preference shocks in region-industry choice. Applying a calibrated China trade shock over 2000–2007, again using Autor, Dorn, and Hanson (2013a), and estimated model elasticities, they find that labor reallocation across sectors and regions is largely complete by thirteen years after shock initiation (i.e., 2013, which is well within our period of analysis). At this horizon, the average change in welfare is 0.20 percent, with a (unweighted) standard deviation across US states of 0.09 percentage points.[50]

48. For an early welfare analysis of China's trade expansion, see Hsieh and Ossa (2016).

49. Lagakos and Waugh (2013) and Hsieh and others (2019) were the first to combine the Eaton and Kortum and Roy models.

50. We thank Lorenzo Caliendo for providing this estimate. Regional labor reallocation does not eliminate variation in welfare impacts across regions because migration costs and region-specific preferences create real adjustment costs.

15985-06a_Autor-5thPgs.indd 430 5/12/22 12:06 PM

In both studies, because goods prices are assumed identical across regions, aggregate welfare impacts are similar.[51] Naturally, regional variation in welfare impacts is larger in Galle, Rodríguez-Clare, and Yi (2017), owing to geographic labor immobility. The two models also differ in how they model labor market frictions, which determine how trade shocks affect the overall employment rate.[52] In Caliendo, Dvorkin, and Parro (2019), nonemployment arises because of an option for home production; in an extended version of Galle, Rodríguez-Clare, and Yi (2017), nonemployment arises because of both a home production option and search and matching frictions in the labor market. Notably, in the latter analysis eliminating search and matching frictions and home production yields little change in the mean and variance of income changes across CZs. Wherever regional variance in income changes is coming from, it does not appear to be labor-market frictions within regions as modeled.

To evaluate quantitative analyses of the China trade shock, we compare changes in the regional dispersion of income per capita in these analyses with reduced-form estimates of the trade-induced change in the variance of income across regions. Because quantitative approaches target the average impact of trade exposure on manufacturing employment in their calibrations, untargeted moments, such as regional dispersion in income changes, are useful for assessing modeling approaches.

In considering regional variation in trade impacts, we abstract from within-region variation in the welfare impacts of trade shocks, which would arise from differential effects of trade exposure on earnings and price levels across households within a CZ. On earnings, Chetverikov, Larsen, and Palmer (2016) and Autor, Dorn, and Hanson (2019) find that lower-wage workers suffered larger percentage declines in earnings in CZs harder hit by the China trade shock. Our focus on personal income per capita aggregates over such worker-level variation. On prices, although the China trade shock appears to have raised consumer purchasing power, Borusyak and Jaravel (2021) document that shares of imports in consumer expenditures

51. Kim and Vogel (2020) provide a related assessment of trade with China, which we discuss in online appendix A.6.

52. In related work, Adão, Arkolakis, and Esposito (2019) modify the baseline framework in Galle, Rodríguez-Clare, and Yi (2017) by making trade between CZs costly and production subject to external economies of scale. If some trade-exposed CZs are proximate to other trade-exposed CZs, the impact of adverse trade shocks will be magnified via localized changes in trade flows, where scale economies may intensify such effects. Their counterfactual analysis implies that the China trade shock reduced average real income by 0.35 percent, with a standard deviation across CZs of 0.56. In online appendix A.5.3, we find weak evidence of gravity-based spillovers of trade shocks across regions.

are flat across the income distribution, which implies that the purchasing power effects of freer trade are distributionally neutral. Relatedly, Bai and Stumpner (2019) and Hottman and Monarch (2020) find that the impact of trade with China on US product prices was similar across income groups (and geographic regions). To a first approximation, we can analyze the impact of the China shock on welfare in the United States as the sum of a price-level effect, which is presumed homogeneous across regions, and an income effect, which varies across regions.[53]

V.B. Relative and Absolute Welfare Effects across Regions

Following Galle, Rodríguez-Clare, and Yi (2017), we can write the average change in welfare for a CZ due to the China trade shock as the induced change in real income,

$$\hat{W}_i = \frac{\hat{Y}_i}{\hat{L}_i} \prod_j \hat{P}_j^{-\beta_j}, \tag{7}$$

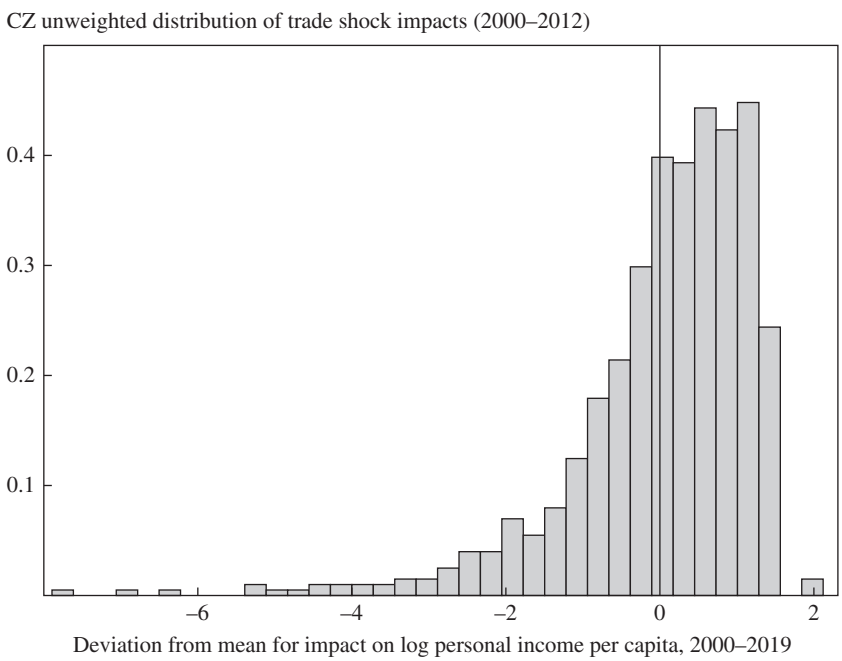
where $\hat{W}_i \equiv W_i'/W_i$ is the trade-induced change in income per capita (Y_{ci}/L_{ci}) in US region i, relative to the change in the national price index, which applies the β_j Cobb-Douglas expenditure shares for products in each sector j to induced changes in the price level P_j for each sector j. Because Galle, Rodríguez-Clare, and Yi (2017) calibrate their model to the 2000–2007 trade shock, they implicitly assume that the long-run impacts of the shock had been realized by 2007. For context, we see in figure 5 that the impact of trade exposure on manufacturing employment over 2001–2007 ($\beta = -1.45$, t-value = –4.56) is approximately 80 percent of that over the 2001–2019 long difference ($\beta = -1.79$, t-value = –5.42). It thus seems reasonable to compare their simulated standard deviation in income changes (0.31) to that estimated for a long-run time change.

As a reduced-form application of the welfare formula in equation (7), we estimate the impact of trade shocks on the variation in average income, Y_{ci}/L_{ci}. Specifically, we evaluate the deviation in changes in income per capita in each CZ relative to the national weighted average,

$$\ln \frac{\hat{Y}_i}{\hat{L}_i} - \sum_l \gamma_l \ln \frac{\hat{Y}_l}{\hat{L}_l} = \beta_1^y \Delta IP_{i\tau}^{cu} - \sum_l \gamma_l \beta_1^y \Delta IP_l^{cu}, \tag{8}$$

53. In online appendix A.6, we discuss results in Feler and Senses (2017) on how trade affects housing values.

Figure 14. Implied Variation in Changes in Personal Income per Capita, 2000–2019



Source: Authors' calculations.

Note: This figure shows a histogram for the welfare change in equation (8), evaluated for the 2000–2012 trade shock–induced change in personal income per capita (based on results in figure 8, panel A), expressed as the deviation in shock impacts from the population-weighted national mean. The impact coefficient used is that for the 2000–2019 time difference ($\beta = -2.66$, t-value = −1.90). The standard deviation in the implied shock impact is 1.22 percentage points.

where γ_l is the share of region l in the US population, β_1^y is the parameter estimate in equation (2) for the log change in personal income per capita, and $IP_{i\tau}^{cu}$ is the trade shock. This estimate of the relative impact leaves unmeasured the shock impact on the overall price level, which we treat as common across regions, as consistent with Caliendo, Dvorkin, and Parro (2019) and Galle, Rodríguez-Clare, and Yi (2017).

In figure 14, we show the estimation-implied variance in equation (8) induced by the 2000–2012 trade shock over the 2000–2019 horizon.[54]

54. The impact coefficient for 2000–2019 in figure 7, panel A, is marginally significant (t-value = −1.90). Impacts are statistically significant for all time differences from 2000–2014 to 2000–2018. Because these coefficients are larger in absolute value than the one for 2000–2019, using the long period difference can be seen as a conservative choice.

The (unweighted) standard deviation in trade shock impacts on personal income per capita across CZs over 2000–2019 is 1.22 percentage points, which far exceeds the cross-CZ income dispersion generated by quantitative models. Online appendix table A4 lists the CZs that are above the 95th percentile of gross income losses—not accounting for any offsetting gains from trade—which is 2.23 percentage points. The number of CZs that suffered net declines in welfare depends, of course, on the magnitude of the gains from trade that potentially offset these gross losses. The high variance of impacts across CZs, relative to apparently small average gains from trade in quantitative analyses, raises the possibility that many CZs suffered net welfare losses. In Caliendo, Dvorkin, and Parro (2019) and Galle, Rodríguez-Clare, and Yi (2017), gains from trade in the aggregate are 0.22 percent or less. When adding this value to the gross losses calculated above, the implied number of CZs experiencing net losses from trade is 223, representing 32.8 percent of the US continental population in 2000. Even if we double these gains to 0.44 percent, 173 CZs, representing 15.9 percent of the US population, would be estimated to have suffered net absolute welfare losses.

The translation of the CZ-level relative effects into absolute level effects hinges on the impact of trade on consumer prices. Jaravel and Sager (2019) estimate the impact of the China trade shock on consumer prices in a reduced-form analysis. Using the 1991–2007 trade shock from Autor and others (2014), they find that a 1.0 percentage point larger increase in import penetration produces a decline in consumer prices of 1.4 percent.[55] Using the mean decadalized change in import penetration over 2000–2012 of 0.89 percentage points (table 1), the implied reduction in consumer prices based on Jaravel and Sager (2019) is 1.25 percent. If the positive impact of trade exposure on consumer prices is indeed this large—over five times the gains from the China trade shock based on quantitative analyses—the number of CZs suffering net losses from trade is smaller but still meaningful. In figure 14, there are 82 CZs, representing 6.3 percent of US population, with declines in personal income per capita of 1.25 percent or more. For these CZs, the net welfare effect of trade with China is negative, even when applying the sizable Jaravel and Sager (2019) price index adjustment. Given the large dispersion of per capita incomes across CZs in figure 14, there would have to be enormous consumer price benefits from trade with China in order to generate net average gains for all US CZs.

55. Dorn and Levell (2021) analyze the impact of trade with China on consumer prices in the United Kingdom and obtain treatment effects that are about half as large as those in Jaravel and Sager (2019)

15985-06a_Autor-5thPgs.indd 434 5/12/22 12:07 PM

VI. Concluding Discussion

Economists have long known that individuals are scarred by job loss. Displaced workers earn significantly less than similar workers who are not displaced, even years after displacement. The China trade shock caused locally concentrated job loss, which led to lasting declines in employment rates and income levels in the most exposed communities. Despite these now well-documented adverse labor market impacts of globalization, there is no consensus about how to remediate such injuries.

The United States does little to protect workers from mass layoff events such as the China trade shock. Although Trade Adjustment Assistance delivers benefits to some workers who have been displaced by trade shocks (Hyman 2018), the program is far too small to have reached most workers who lost their jobs because of import competition in the 1990s and 2000s. A further limitation of Trade Adjustment Assistance is that it conditions assistance on the cause of job loss, that is, trade. Presumably, job loss is equally scarring no matter whether the underlying cause of displacement is import competition, technological change (Autor and Dorn 2013; Acemoglu and Restrepo 2020), government regulation (Walker 2013), or some other factor. Policy failures in this domain are far from innocuous. The political gains of US and European nationalist-populists, for instance, have been greater in regions that have suffered larger trade-induced employment declines (Colantone and Stanig 2018a, 2018b; Autor and others 2020a; Rodrik 2020).

The favored solution of populist politicians to regional distress is to raise import barriers and block immigration. Indeed, the Donald J. Trump administration cited the adverse impacts of the China trade shock to justify taking aggressive trade action against China (Redding 2020). The subsequent US-China trade war succeeded in elevating US product prices (Amiti, Redding, and Weinstein 2019; Fajgelbaum and others 2020; Cavallo and others 2021) but not in expanding employment in import-protected sectors (Flaaen and Pierce 2019). We are aware of no research that would justify ex post protectionist trade measures as a means of helping workers hurt by past import competition.

Recent literature suggests that fostering employment growth in regions beset by chronic joblessness may help workers hurt by persistent negative local labor demand shocks.[56] Austin, Glaeser, and Summers (2018) find that employment impacts of labor demand shocks are larger in local markets in

56. On optimal spatial policies, see Fajgelbaum and Gaubert (2020) and Gaubert, Kline, and Yagan (2021).

which joblessness was initially high. Bartik (2020) argues further that job growth in distressed regions is especially beneficial for low-wage workers. Absent success in helping regions left behind by globalization, one may expect continued popular support for political platforms that disparage foreign trade, despite the apparent inefficacy of these platforms to date.

Because of the distinctiveness of the China trade shock—in terms of its enduring magnitude and concentration in place and time—some see it as holding few lessons for policymakers. Such characterizations are mistaken. Large, localized contractions in labor demand—for example, trade and technology shocks to autos and steel in the 1970s, price shocks to coal in the 1980s, and ongoing automation and roboticization—are unfortunately common events. The global economy is now witnessing another round of localized shocks, arising from the energy transition. If economies succeed in sharply reducing reliance on fossil fuels, demand will plummet for the workers who mine coal, drill for oil and gas, refine petroleum, and transport and distribute these outputs (Popp and others 2021). Industries that are energy intensive in production and that located their operations close to carbon-based energy sources may also experience significant disruptions. Although increased demand for renewable energy will surely create new jobs in harnessing solar, wind, and hydro power, these jobs may be located far from existing carbon-based production facilities. New energy-intensive industries, such as data centers, may choose to locate near renewable energy providers, possibly further disadvantaging those working in the carbon-based economy. We are now armed with knowledge both of the consequences of localized job loss and the likely location of future worker displacement. This should enable policymakers to prepare for assisting workers, as carbon-based jobs disappear, and regions, as they reorient toward less carbon-intensive activities. Without such preparation, the energy transition may add to the painful history of regionalized job loss.

It is now clear that the China trade shock as we understand it appears to have stopped intensifying a decade ago. The China shock of the 1990s and 2000s was about China's onetime transition to a market economy. This transition, which affected US industries that were late in their economic life cycles, was in some sense forecastable based on market fundamentals, both in China and the United States. The next China shock may be more likely to be triggered by industrial policy, for example, if China makes good on its promise to support advanced technology industries. Such a shock may depend less on market fundamentals and more on China's future interventionist policy choices—and the reactions of the United States and other countries to these choices—making its dimensions difficult to foresee.

15985-06a_Autor-5thPgs.indd 436 5/12/22 12:07 PM

Because of China's immense scale and willingness to enact sweeping policy changes in short order, future China shocks may again have profound albeit unpredictable consequences.

ACKNOWLEDGMENTS For helpful comments, we thank Loren Brandt, Lorenzo Caliendo, Federico Esposito, Bruno Ferman, Stephen Redding, Andrés Rodríguez-Clare, Katheryn Russ, Jonathan Vogel, and, especially, Marianne Bertrand, Edward Glaeser, and James Stock. Anne Beck, Peter Bruno, Brendan Chapuis, Rebecca Jackson, René Livas, Sophia Mo, and Savannah Noray provided excellent research assistance. Autor thanks the following for financial support: Smith Richardson Foundation (#20202252), Accenture LLP (#027843-0001), and Andrew Carnegie Fellowship (G-F-19-56882). Dorn acknowledges support by the University of Zurich's Research Priority Program Equality of Opportunity.

References

Acemoglu, Daron, David H. Autor, David Dorn, Gordon H. Hanson, and Brendan Price. 2016. "Import Competition and the Great US Employment Sag of the 2000s." *Journal of Labor Economics* 34, no. S1: S141–98.

Acemoglu, Daron, and Pascual Restrepo. 2020. "Robots and Jobs: Evidence from US Labor Markets." *Journal of Political Economy* 128, no. 6: 2188–244.

Adão, Rodrigo, Costas Arkolakis, and Federico Esposito. 2019. "Spatial Linkages, Global Shocks, and Local Labor Markets: Theory and Evidence." Working Paper 2019-10. Chicago: University of Chicago, Becker Friedman Institute for Economics.

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-Share Designs: Theory and Inference." *Quarterly Journal of Economics* 134, no. 4: 1949–2010.

Aghion, Philippe, Antonin Bergeaud, Matthieu Lequien, Marc Melitz, and Thomas Zuber. 2021. "Opposing Firm-Level Responses to the China Shock: Horizontal Competition versus Vertical Relationships?" Working Paper 29196. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w29196.

Allcott, Hunt, and Daniel Keniston. 2017. "Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America." *Review of Economic Studies* 85, no. 2: 695–731.

Amior, Michael, and Alan Manning. 2018. "The Persistence of Local Joblessness." *American Economic Review* 108, no. 7: 1942–70.

Amiti, Mary, Stephen J. Redding, and David E. Weinstein. 2019. "The Impact of the 2018 Tariffs on Prices and Welfare." *Journal of Economic Perspectives* 33, no. 4: 187–210.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103, no. 484: 1481–95.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New Trade Models, Same Old Gains?" *American Economic Review* 102, no. 1: 94–130.

Asquith, Brian, Sanjana Goswami, David Neumark, and Antonio Rodriguez-Lopez. 2019. "U.S. Job Flows and the China Shock." *Journal of International Economics* 118:123–37.

Austin, Benjamin, Edward L. Glaeser, and Lawrence Summers. 2018. "Jobs for the Heartland: Place-Based Policies in 21st-Century America." *Brookings Papers on Economic Activity*, Spring, 151–232.

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

Autor, David H., David Dorn, and Gordon H. Hanson. 2013a. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103, no. 6: 2121–68.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013b. "The Geography of Trade and Technology Shocks in the United States." *American Economic Review* 103, no. 3: 220–225.

Autor, David H., David Dorn, and Gordon H. Hanson. 2016. "The China Shock: Learning from Labor Market Adjustment to Large Changes in Trade." *Annual Review of Economics* 8, no. 1: 205–40.

Autor, David H., David Dorn, and Gordon H. Hanson. 2019. "When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men." *American Economic Review: Insights* 1, no. 2: 161–78.

Autor, David H., David Dorn, Gordon H. Hanson, and Kaveh Majlesi. 2020a. "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure." *American Economic Review* 110, no. 10: 3139–83.

Autor, David H., David Dorn, Gordon H. Hanson, Gary Pisano, and Pian Shu. 2020b. "Foreign Competition and Domestic Innovation: Evidence from US Patents." *American Economic Review: Insights* 2, no. 3: 357–74.

Autor, David H., David Dorn, Gordon H. Hanson, and Jae Song. 2014. "Trade Adjustment: Worker-Level Evidence." *Quarterly Journal of Economics* 129, no. 4: 1799–860.

Bai, Liang, and Sebastian Stumpner. 2019. "Estimating US Consumer Gains from Chinese Imports." *American Economic Review: Insights* 1, no. 2: 209–24.

Bai, Xue, Kala Krishna, and Hong Ma. 2017. "How You Export Matters: Export Mode, Learning and Productivity in China." *Journal of International Economics* 104:122–37.

Balassa, Bela. 1965. "Trade Liberalisation and 'Revealed' Comparative Advantage." *Manchester School* 33, no. 2: 99–123.

Bartik, Alexander W., Janet Currie, Michael Greenstone, and Christopher R. Knittel. 2019. "The Local Economic and Welfare Consequences of Hydraulic Fracturing." *American Economic Journal: Applied Economics* 11, no. 4: 105–55.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, Mich.: W. E. Upjohn Institute.

Bartik, Timothy J. 2020. "Should Place-Based Jobs Policies Be Used to Help Distressed Communities?" *Journal of Economic Perspectives* 34, no. 3: 99–127.

Batistich, Mary Kate, and Timothy N. Bond. 2019. "Stalled Racial Progress and Japanese Trade in the 1970s and 1980s." Working Paper 12133. Bonn: IZA Institute of Labor Economics.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. "Productive Cities: Sorting, Selection, and Agglomeration." *Journal of Political Economy* 122, no. 3: 507–53.

Beraja, Martin, Erik Hurst, and Juan Ospina. 2019. "The Aggregate Implications of Regional Business Cycles." *Econometrica* 87, no. 6: 1789–833.

Bergsten, C. Fred, and Joseph E. Gagnon. 2017. *Currency Conflict and Trade Policy: A New Strategy for the United States*. Washington: Peterson Institute for International Economics.

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

Black, Dan, Kermit Daniel, and Seth Sanders. 2002. "The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust." *American Economic Review* 92, no. 1: 27–50.

Black, Dan, Terra McKinnish, and Seth Sanders. 2005. "The Economic Impact of the Coal Boom and Bust." *Economic Journal* 115, no. 503: 449–76.

Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity*, no. 1: 1–75.

Bloom, Nicholas, Kyle Handley, Andre Kurman, and Phillip Luck. 2019. "The Impact of Chinese Trade on U.S. Employment: The Good, the Bad, and the Debatable." https://nbloom.people.stanford.edu/sites/g/files/sbiybj4746/f/bhkl_posted_draft.pdf.

Boehm, Christoph E., Aaron Flaaen, and Nitya Pandalai-Nayar. 2020. "Multinationals, Offshoring, and the Decline of US Manufacturing." *Journal of International Economics* 127:103391.

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

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-Experimental Shift-Share Research Designs." *Review of Economic Studies* 89, no. 1: 181–213.

Borusyak, Kirill, and Xavier Jaravel. 2021. "The Distributional Effects of Trade: Theory and Evidence from the United States." Working Paper 28957. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28957.

Bound, John, and Harry J. Holzer. 2000. "Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s." *Journal of Labor Economics* 18, no. 1: 20–54.

Brandt, Loren, and Kevin Lim. 2020. "Accounting for Chinese Exports." Working Paper. https://www.economics.utoronto.ca/public/workingPapers/tecipa-680.pdf.

Brandt, Loren, John Litwack, Elitza Mileva, Luhang Wang, Yifan Zhang, and Luan Zhao. 2020. "China's Productivity Slowdown and Future Growth Potential." Working Paper 9298. Washington: World Bank Group. https://openknowledge.worldbank.org/handle/10986/33993.

Brandt, Loren, and Peter M. Morrow. 2017. "Tariffs and the Organization of Trade in China." *Journal of International Economics* 104:85–103.

Brandt, Loren, Trevor Tombe, and Xiaodong Zhu. 2013. "Factor Market Distortions across Time, Space and Sectors in China." *Review of Economic Dynamics* 16, no. 1: 39–58.

Brandt, Loren, Johannes Van Biesebroeck, Luhang Wang, and Yifan Zhang. 2017. "WTO Accession and Performance of Chinese Manufacturing Firms." *American Economic Review* 107, no. 9: 2784–820.

15985-06a_Autor-5thPgs.indd 440 5/12/22 12:07 PM

Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang. 2012. "Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing." *Journal of Development Economics* 97, no. 2: 339–51.

Cadena, Brian C., and Brian K. Kovak. 2016. "Immigrants Equilibrate Local Labor Markets: Evidence from the Great Recession." *American Economic Journal: Applied Economics* 8, no. 1: 257–90.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro. 2019. "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock." *Econometrica* 87, no. 3: 741–835.

Cavallo, Alberto, Gita Gopinath, Brent Neiman, and Jenny Tang. 2021. "Tariff Pass-Through at the Border and at the Store: Evidence from US Trade Policy." *American Economic Review: Insights* 3, no. 1: 19–34.

Chan, Jeff. 2019. "Labour Market Characteristics and Surviving Import Shocks." *World Economy* 42, no. 5: 1288–315.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2016. "The Masking of the Decline in Manufacturing Employment by the Housing Bubble." *Journal of Economic Perspectives* 30, no. 2: 179–200.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2018. "Housing Booms and Busts, Labor Market Opportunities, and College Attendance." *American Economic Review* 108, no. 10: 2947–94.

Charles, Kerwin Kofi, Erik Hurst, and M. Schwartz. 2019. "The Transformation of Manufacturing and the Decline in US Employment." *NBER Macroeconomics Annual* 33:307–72.

Chetverikov, Denis, Bradley Larsen, and Christopher Palmer. 2016. "IV Quantile Regression for Group-Level Treatments, with an Application to the Distributional Effects of Trade." *Econometrica* 84, no. 2: 809–33.

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. 2020. "Regional Data in Macroeconomics: Some Advice for Practitioners." *Journal of Economic Dynamics and Control* 115:103875.

Clay, Karen, and Margarita Portnykh. 2018. "The Short-Run and Long-Run Effects of Resources on Economic Outcomes: Evidence from the United States 1936–2015." Working Paper 24695. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24695.

Colantone, Italo, and Piero Stanig. 2018a. "Global Competition and Brexit." *American Political Science Review* 112, no. 2: 201–18.

Colantone, Italo, and Piero Stanig. 2018b. "The Trade Origins of Economic Nationalism: Import Competition and Voting Behavior in Western Europe." *American Journal of Political Science* 62, no. 4: 936–53.

Dauth, Wolfgang, Sebastian Findeisen, and Jens Suedekum. 2014. "The Rise of the East and the Far East: German Labor Markets and Trade Integration." *Journal of the European Economic Association* 12, no. 6: 1643–75.

Davis, Steven J., and John C. Haltiwanger. 2019. "Dynamism Diminished: The Role of Housing Markets and Credit Conditions." Working Paper 25466. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25466.

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

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000." *American Economic Review* 106, no. 3: 479–524.

Dix-Carneiro, Rafael, and Brian K. Kovak. 2017. "Trade Liberalization and Regional Dynamics." *American Economic Review* 107, no. 10: 2908–46.

Dix-Carneiro, Rafael, João Paulo Pessoa, Ricardo M. Reyes-Heroles, and Sharon Traiberman. 2021. "Globalization, Trade Imbalances and Labor Market Adjustment." Working Paper 28315. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28315.

Dorn, David. 2009. "Essays on Inequality, Spatial Interaction, and the Demand for Skills." Diss., University of St. Gallen. https://www.e-helvetica.nb.admin.ch/api/download/urn%3Anbn%3Ach%3Abel-152672%3Adis3613.pdf/dis3613.pdf.

Dorn, David, and Peter Levell. 2021. "Trade and Inequality in Europe and the US." Working Paper 14914. Bonn: IZA Institute of Labor Economics.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70, no. 5: 1741–79.

Economist. 2015. "Made in China?" March 12. https://www.economist.com/leaders/2015/03/12/made-in-china.

Economist. 2016. "Good Afternoon, Vietnam." August 6. https://www.economist.com/finance-and-economics/2016/08/06/good-afternoon-vietnam.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr. 2010. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review* 100, no. 3: 1195–213.

Eriksson, Katherine, Katheryn Russ, Jay C. Shambaugh, and Minfei Xu. 2019. "Trade Shocks and the Shifting Landscape of US Manufacturing." Working Paper 25646. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25646.

Faber, Marius, Andres Sarto, and Marco Tabellini. 2019. "The Impact of Technology and Trade on Migration: Evidence from the US." Working Paper 20-071. Boston: Harvard Business School. https://hbswk.hbs.edu/item/the-impact-of-technology-and-trade-on-migration-evidence-from-the-us.

Fajgelbaum, Pablo D., and Cecile Gaubert. 2020. "Optimal Spatial Policies, Geography, and Sorting." *Quarterly Journal of Economics* 135, no. 2: 959–1036.

Fajgelbaum, Pablo D., Pinelopi K. Goldberg, Patrick J. Kennedy, and Amit K. Khandelwal. 2020. "The Return to Protectionism." *Quarterly Journal of Economics* 135, no. 1: 1–55.

Feenstra, Robert C. 2016. *Advanced International Trade: Theory and Evidence*. Princeton, N.J.: Princeton University Press.

Feenstra, Robert C., and Gordon H. Hanson. 2005. "Ownership and Control in Outsourcing to China: Estimating the Property-Rights Theory of the Firm." *Quarterly Journal of Economics* 120, no. 2: 729–61.

Feenstra, Robert C., Hong Ma, and Yuan Xu. 2019. "US Exports and Employment." *Journal of International Economics* 120:46–58.

Feler, Leo, and Mine Z. Senses. 2017. "Trade Shocks and the Provision of Local Public Goods." *American Economic Journal: Economic Policy* 9, no. 4: 101–43.

Ferman, Bruno. 2019. "Assessing Inference Methods." ArXiv:1912.08772. Ithaca, N.Y.: Cornell University.

Feyrer, James, Erin T. Mansur, and Bruce Sacerdote. 2017. "Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution." *American Economic Review* 107, no. 4: 1313–34.

Feyrer, James, Bruce Sacerdote, and Ariel Dora Stern. 2007. "Did the Rust Belt Become Shiny? A Study of Cities and Counties That Lost Steel and Auto Jobs in the 1980s." *Brookings-Wharton Papers on Urban Affairs*, 41–102.

Flaaen, Aaron, and Justin R. Pierce. 2019. "Disentangling the Effects of the 2018–2019 Tariffs on a Globally Connected US Manufacturing Sector." Finance and Economics Discussion Series 2019-086. Washington: Board of Governors of the Federal Reserve System. https://doi.org/10.17016/FEDS.2019.086.

Fort, Teresa C., Justin R. Pierce, and Peter K. Schott. 2018. "New Perspectives on the Decline of US Manufacturing Employment." *Journal of Economic Perspectives* 32, no. 2: 47–72.

Galle, Simon, Andrés Rodríguez-Clare, and Moises Yi. 2017. "Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade." Working Paper 23737. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w23737.

Gaubert, Cecile, Patrick M. Kline, and Danny Yagan. 2021. "Place-Based Redistribution." Working Paper 28337. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w28337.

Glaeser, Edward L., and Joseph Gyourko. 2005. "Urban Decline and Durable Housing." *Journal of Political Economy* 113, no. 2: 345–75.

Glaeser, Edward L., Joseph Gyourko, and Raven Saks. 2005. "Why Is Manhattan So Expensive? Regulation and the Rise in Housing Prices." *Journal of Law and Economics* 48, no. 2: 331–70.

Glaeser, Edward L., Jose A. Scheinkman, and Andrei Shleifer. 1995. "Economic Growth in a Cross-Section of Cities." *Journal of Monetary Economics* 36, no. 1: 117–43.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review* 110, no. 8: 2586–624.

Greenland, Andrew, and John Lopresti. 2016. "Import Exposure and Human Capital Adjustment: Evidence from the U.S." *Journal of International Economics* 100:50–60.

Greenland, Andrew, John Lopresti, and Peter McHenry. 2019. "Import Competition and Internal Migration." *Review of Economics and Statistics* 101, no. 1: 44–59.

Grossman, Gene M. 1982. "The Employment and Wage Effects of Import Competition in the United States." Working Paper 1041. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w1041.

Guren, Adam, Alisdair McKay, Emi Nakamura, and Jón Steinsson. 2021. "What Do We Learn from Cross-Regional Empirical Estimates in Macroeconomics?" *NBER Macroeconomics Annual* 35:175–223.

Handley, Kyle, and Nuno Limão. 2017. "Policy Uncertainty, Trade, and Welfare: Theory and Evidence for China and the United States." *American Economic Review* 107, no. 9: 2731–83.

Hanson, Gordon H. 2020. "Who Will Fill China's Shoes? The Global Evolution of Labor-Intensive Manufacturing." *East Asian Economic Review* 24, no. 4: 313–36.

Hanson, Gordon H., Nelson Lind, and Marc-Andreas Muendler. 2015. "The Dynamics of Comparative Advantage." Working Paper 21753. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w21753.

Heath, Rachel, and A. Mushfiq Mobarak. 2015. "Manufacturing Growth and the Lives of Bangladeshi Women." *Journal of Development Economics* 115:1–15.

Heckman, James J., Lance Lochner, and Christopher Taber. 1998. "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents." *Review of Economic Dynamics* 1, no. 1: 1–58.

Helpman, Elhanan. 2018. *Globalization and Inequality*. Cambridge, Mass.: Harvard University Press.

Holmes, Thomas J., and John J. Stevens. 2014. "An Alternative Theory of the Plant Size Distribution, with Geography and Intra- and International Trade." *Journal of Political Economy* 122, no. 2: 369–421.

Hombert, Johan, and Adrien Matray. 2018. "Can Innovation Help U.S. Manufacturing Firms Escape Import Competition from China?" *Journal of Finance* 73, no. 5: 2003–39.

Hottman, Colin J., and Ryan Monarch. 2020. "A Matter of Taste: Estimating Import Price Inflation across U.S. Income Groups." *Journal of International Economics* 127:103382.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. 2019. "The Allocation of Talent and U.S. Economic Growth." *Econometrica* 87, no. 5: 1439–74.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124, no. 4: 1403–48.

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation." *American Economic Journal: Macroeconomics* 11, no. 2: 1–39.

Hsieh, Chang-Tai, and Ralph Ossa. 2016. "A Global View of Productivity Growth in China." *Journal of International Economics* 102:209–24.

Hsieh, Chang-Tai, and Zheng Michael Song. 2015. "Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China." *Brookings Papers on Economic Activity*, Spring, 295–366.

Hu, Albert Guangzhou, and Gary H. Jefferson. 2009. "A Great Wall of Patents: What Is Behind China's Recent Patent Explosion?" *Journal of Development Economics* 90, no. 1: 57–68.

Huckfeldt, Christopher. 2021. "Understanding the Scarring Effect of Recessions." *American Economic Review*, forthcoming.

Hyman, Benjamin. 2018. "Can Displaced Labor Be Retrained? Evidence from Quasi-Random Assignment to Trade Adjustment Assistance." Working Paper. Social Science Research Network, April 20. https://ssrn.com/abstract=3155386.

Jacobsen, Grant D., and Dominic P. Parker. 2014. "The Economic Aftermath of Resource Booms: Evidence from Boomtowns in the American West." *Economic Journal* 126, no. 593: 1092–128.

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

Jaravel, Xavier, and Erick Sager. 2019. "What Are the Price Effects of Trade? Evidence from the US and Implications for Quantitative Trade Models." CEP Discussion Paper 1642. London: Center for Economic Performance.

Kearney, Melissa S., and Riley Wilson. 2018. "Male Earnings, Marriageable Men, and Nonmarital Fertility: Evidence from the Fracking Boom." *Review of Economics and Statistics* 100, no. 4: 678–90.

Kee, Hiau Looi, and Heiwai Tang. 2016. "Domestic Value Added in Exports: Theory and Firm Evidence from China." *American Economic Review* 106, no. 6: 1402–36.

Khandelwal, Amit K., Peter K. Schott, and Shang-Jin Wei. 2013. "Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters." *American Economic Review* 103, no. 6: 2169–95.

Kim, Ryan, and Jonathan Vogel. 2020. "Trade and Welfare (across Local Labor Markets)." Working Paper 27133. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27133.

Kim, Sukkoo. 1995. "Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860–1987." *Quarterly Journal of Economics* 110, no. 4: 881–908.

Kletzer, Lori G. 2001. *Job Loss from Imports: Measuring the Costs*. Washington: Peterson Institute for International Economics.

Koopman, Robert, Zhi Wang, and Shang-Jin Wei. 2012. "Estimating Domestic Content in Exports When Processing Trade Is Pervasive." *Journal of Development Economics* 99, no. 1: 178–89.

Lagakos, David, and Michael E. Waugh. 2013. "Selection, Agriculture, and Cross-Country Productivity Differences." *American Economic Review* 103, no. 2: 948–80.

Lardy, Nicholas R. 2019. *The State Strikes Back: The End of Economic Reform in China?* Washington: Peterson Institute for International Economics.

Li, Lei. 2020. "Skill-Biased Imports, Human Capital Accumulation, and the Allocation of Talent." CRC TR 224 Discussion Paper. https://www.kcg-kiel.org/wp-content/uploads/2020/06/LI_KCG-Seminar-on-260620.pdf.

Liu, Chen, and Xiao Ma. 2020. "China's Export Surge and the New Margins of Trade." Working Paper. https://acsweb.ucsd.edu/~xim032/pdfs/China_shock_draft.pdf

Ma, Xiao. 2020. "College Expansion, Trade, and Innovation: Evidence from China." Working Paper. https://acsweb.ucsd.edu/~xim032/pdfs/college_inno.pdf.

Maneschi, Andrea. 2008. "How Would David Ricardo Have Taught the Principle of Comparative Advantage?" *Southern Economic Journal* 74, no. 4: 1167–76.

Mian, Atif, and Amir Sufi. 2014. "What Explains the 2007–2009 Drop in Employment?" *Econometrica* 82, no. 6: 2197–223.

Michaels, Guy. 2008. "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System." *Review of Economics and Statistics* 90, no. 4: 683–701.

Michaels, Guy. 2010. "The Long Term Consequences of Resource-Based Specialisation." *Economic Journal* 121, no. 551: 31–57.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. 2018. "Commuting, Migration, and Local Employment Elasticities." *American Economic Review* 108, no. 12: 3855–90.

Naughton, Barry J. 2006. *The Chinese Economy: Transitions and Growth*. Cambridge, Mass.: MIT Press.

Naughton, Barry J. 2016. "Inside and Outside: The Modernized Hierarchy That Runs China." *Journal of Comparative Economics* 44:404–415.

Notowidigdo, Matthew J. 2020. "The Incidence of Local Labor Demand Shocks." *Journal of Labor Economics* 38, no. 3: 687–725.

Pierce, Justin R., and Peter K. Schott. 2016. "The Surprisingly Swift Decline of US Manufacturing Employment." *American Economic Review* 106, no. 7: 1632–62.

Pierce, Justin R., and Peter K. Schott. 2020. "Trade Liberalization and Mortality: Evidence from US Counties." *American Economic Review: Insights* 2, no. 1: 47–64.

Popp, David, Francesco Vona, Giovanni Marin, and Ziqiao Chen. 2021. "The Employment Impact of a Green Fiscal Push: Evidence from the American Recovery and Reinvestment Act." In the present volume of *Brookings Papers on Economic Activity*.

Redding, Stephen J. 2020. "Trade and Geography." Working Paper 27821. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w27821.

Ricardo, David. 1815. *An Essay on the Influence of a Low Price of Corn on the Profits of Stock, with Remarks on Mr. Malthus' Two Last Publications*. Oxford University.

Rodríguez-Clare, Andres, Mauricio Ulate, and Jose P. Vásquez. 2020. "New-Keynesian Trade: Understanding the Employment and Welfare Effects of Sector-Level Shocks." Working Paper 2020-32. Federal Reserve Bank of San Francisco. https://www.frbsf.org/economic-research/files/wp2020-32.pdf.

15985-06a_Autor-5thPgs.indd 446 5/12/22 12:07 PM

Rodrik, Dani. 2020. "Why Does Globalization Fuel Populism? Economics, Culture, and the Rise of Right-Wing Populism." Working Paper 27526. Cambridge, Mass.: National Bureau of Economic Research.

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3, no. 2: 135–46.

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti. 2011. "Growing Like China." *American Economic Review* 101, no. 1: 196–233.

Tolbert, Charles M., and Molly Sizer. 1996. "US Commuting Zones and Labor Market Areas: A 1990 Update." Staff Paper AGES-9614. Washington: US Department of Agriculture. https://usa.ipums.org/usa/resources/volii/cmz90.pdf.

Vogel, Ezra F. 2011. *Deng Xiaoping and the Transformation of China*. Cambridge, Mass.: Harvard University Press.

Walker, W. Reed. 2013. "The Transitional Costs of Sectoral Reallocation: Evidence from the Clean Air Act and the Workforce." *Quarterly Journal of Economics* 128, no. 4: 1787–835.

Wei, Shang-Jin, Zhuan Xie, and Xiaobo Zhang. 2017. "From 'Made in China' to 'Innovated in China': Necessity, Prospect, and Challenges." *Journal of Economic Perspectives* 31, no. 1: 49–70.

Wilson, William Julius. 1996. *When Work Disappears: The World of the New Urban Poor*. New York: Alfred A. Knopf; Vintage Books, 1997.

Wolf, Christian K. 2019. "The Missing Intercept in Cross-Regional Regressions." Working Paper. https://scholar.princeton.edu/sites/default/files/ckwolf/files/missing_intercept.pdf.

Wood, Adrian. 1995. "How Trade Hurt Unskilled Workers." *Journal of Economic Perspectives* 9, no. 3: 57–80.

Wood, Adrian. 2018. "The 1990s Trade and Wages Debate in Retrospect." *World Economy* 41, no. 4: 975–99.

Yagan, Danny. 2019. "Employment Hysteresis from the Great Recession." *Journal of Political Economy* 127, no. 5: 2505–58.

Yu, Miaojie. 2015. "Processing Trade, Tariff Reductions, and Firm Productivity: Evidence from Chinese Firms." *Economic Journal* 125, no. 585: 943–88.

Comments and Discussion

COMMENT BY

MARIANNE BERTRAND Autor, Dorn, and Hanson build on their highly influential prior work on the China shock to examine longer-term adjustments in the local labor markets most exposed to Chinese imports during the 2000s. Their main takeaway is that while Chinese imports reached a plateau in 2010, adverse impacts on manufacturing employment, overall employment-population ratios, and income per capita were still present in the more negatively exposed areas as of 2019. Expected labor market adjustment via out-migration to areas with better job opportunities broadly failed to occur. Looking at heterogeneity across commuting zones (CZs), the authors show that the persistence of adverse outcomes is more acute in areas that started with high levels of industrial specialization as well as with a less-educated workforce. Using some of the most generous estimates of the consumer benefits of rising trade with China, the authors conclude that more than 5 percent of the US population may have suffered absolute declines in average real incomes due to China's entry into the World Trade Organization (WTO).

AGGREGATE TIME SERIES PERSPECTIVE The authors, both in the current paper and in their prior work, have been mainly concerned with the regional variation in exposure to the China trade shock and have mostly stayed away from discussing the aggregate implications of China's entry into the WTO for the US labor market. Yet their findings have at times been hijacked to feed a populist backlash against free trade that has put the blame for the decline of US manufacturing employment on a rape by China.[1] The authors' work has inspired an active and growing literature that attempts

1. "Trump Accuses China of 'Raping' US with Unfair Trade Policy," BBC News, May 2, 2016, https://www.bbc.com/news/election-us-2016-36185012.

15985-06b_AutorCom&GD-3rdPgs.indd 448 5/12/22 12:06 PM

to provide answers to the aggregate impact question, a question that is less well suited to a cross-regional, reduced-form identification strategy. It was worthwhile for me, as a discussant who is an outsider to this literature, to start with a review of the aggregate time series evidence on manufacturing employment in the United States.

Bureau of Labor Statistics data reveal a linear decline in the share of manufacturing employment from about the mid-1960s to the early 2010s, with a stabilization since.[2] Economic recessions are typically associated with somewhat faster decline in the manufacturing share (with the COVID-19 recession a likely outlier), and the 2000s were characterized by two such recessions. Beyond this, it is hard to discern anything off-trend about the first decade of the new millennium, which coincided with China's entry into the WTO.

As pointed out to me by the authors, a steady downward trend in the manufacturing share of overall employment may mask more complex dynamics when it comes to the number of manufacturing versus non-manufacturing jobs in the economy. Indeed, in levels, there is much clearer evidence of faster drops in the number of manufacturing jobs in the 2000s, compared to a steady level from the mid-1960s to the late 1990s.[3] Yet, even in the level series, a direct link between the China shock and these aggregate manufacturing employment losses seems somewhat tenuous. In particular, the manufacturing job losses are concentrated during the 2000–2001 recession and during the Great Recession. The number of people employed in manufacturing is pretty constant from 2003 to the beginning of the Great Recession. This is in contrast with the very smooth increases in China's share of world manufacturing exports, as seen in figure 1 in the paper, and in China's share of US imports, as seen in figure 3. Furthermore, a large share of the employment losses during the first of the two dips occurred before December 2001, which marked China's entry into the WTO. It is possible, as Pierce and Schott (2016) have argued, that the China shock truly began more than a year earlier, when the US Congress granted permanent normal trade relations to China in October 2000. Nevertheless, the contrast between the smooth increase in Chinese imports during the 2000s and the more business-cyclical changes in manufacturing employment over the same period raises questions as to how first-order

2. FRED Economic Data, "All Employees, Manufacturing/All Employees, Total Nonfarm," https://fred.stlouisfed.org/graph/?g=cAYh.

3. FRED Economic Data, "All Employees, Manufacturing," https://fred.stlouisfed.org/series/MANEMP.

the China shock might have been in explaining the overall decline in US manufacturing since the beginning of the new millennium. To be clear, this does not take away from the authors' core interest in how local experiences of extreme exposure to Chinese imports may have left some communities economically battered for the longer run.

MEASUREMENT OF THE CHINA SHOCK Building on Autor, Dorn, and Hanson (2013) and Acemoglu and others (2016), the authors measure exposure to the China shock at the US CZ level. Exposure as of year *t* is defined using a shift-share approach: average change in Chinese import penetration across industries between year *t* and a baseline year (in most cases 2000), weighted by industry shares in the CZ employment in the baseline year.

While this is a highly reasonable measure of Chinese import competition in a given CZ, it is much less clear to me why it would be the best proxy for an area's overall exposure to the China trade shock. First, this measure excludes US exports to China, which have also grown since China's entry into the WTO. The authors argue that this exclusion is justifiable because US imports from China dwarf US exports to China: they mention that the 2000–2012 increase in US manufacturing imports from China was 4.1 times the increase in US manufacturing exports to China. Yet one may counter that a quarter of a massive number is still a massive number. Furthermore, unlike US imports from China, which have plateaued since 2010, US exports to China have continued to increase; this would further speak in favor of including them in an analysis of the long-term effects of the China shock.

Also, and as importantly, a sole focus on direct import competition from China abstracts away from a richer supply chain perspective and the upstream and downstream channels via which the trade shock may have had an impact on local economies. For example, Wang and others (2018) use a methodology at first blush similar to Autor, Dorn, and Hanson (2013) and suggest that the employment benefits from being able to use (cheaper) imported intermediate goods, that is, downstream effects, outweigh the negative effects of direct import competition and of being upstream of US firms that directly compete with China.

In summary, it is unclear, except maybe for reasons of path dependence in research, as to why the authors stick exclusively with one channel of impact for the China shock, while the data exist to consider the other channels and the identification strategy appears generalizable to these other channels.

IDENTIFICATION There are several empirical challenges the authors face while trying to assess the causal impact of differential geographical exposure to Chinese imports. The first challenge is with regard to the endogeneity

of US imports from China to domestic demand conditions. To address this concern and identify the foreign supply–driven component of this change in import penetration, the authors instrument the US import exposure measure with non-US exposure, using industry-level growth of Chinese exports to other developed countries.

A second challenge, which the instrumentation strategy cannot address, is that the low-skill industrial sectors that were most exposed to direct competition from China throughout the developed countries might be the same sectors that were also exposed to skilled-biased technological change and the adoption of other labor-saving technologies. Here, the authors' approach is to control for as much as is measurable and relevant, which the authors do by including controls for baseline specialization in occupations according to their routine-task intensity and offshorability, following Autor and Dorn (2013). It is unclear whether these controls can fully capture these technological forces and there is little in the analysis to assess the sensitivity of the results to other proxies for these forces (such as capital investment measures). Furthermore, it is unclear how to think about a control such as offshorability. One could argue that this is overcontrolling to the extent that some US firms may have responded to the shock via offshoring some of their operations to China, as suggested by Pierce and Schott (2016).

Other baseline controls are meant to account for other sources of variation across CZs that might be correlated with the import exposure measure and could explain changes in employment, labor force participation, or the take-up of government transfers over the period under study. Again here, little sensitivity analysis is reported on the list of controls included. For example, one excluded control I was curious about is the baseline share of male employment in the CZ. The male employment share might be correlated with exposure to the China shock and could be a nontrivial force in the longer-term trends the authors document, given the secular decline in US prime-age male employment and labor force participation over the last fifty years, especially among less-educated men (Binder and Bound 2019).[4] More generally, given the non-randomness of which CZs were most exposed to Chinese imports, it would be valuable to discuss in more detail how the baseline controls were selected and how sensitive the results are to different choices.

4. Share of women (and hence men) in CZ employment is included as a control in the analysis of labor market adjustment to the decline of coal. Instead, as far as I understand it, the analysis of labor market adjustment to the China shock includes the fraction of working-age women who are employed.

The authors further argue that none of the forces or factors above would have changed as abruptly as the China shock. This is a fair point. One easy way to further substantiate this would be to report pre-trends. The figures presented in the paper do not allow the reader to assess these pre-trends, as they only display effects starting in 2002. The authors could easily start these figures, say, a decade earlier, to help the reader assess whether or not the CZs experiencing the greatest impact from the China shock were (or were not) already differentially trending prior to 2002. This is the approach they follow for the analysis of the long-term impacts of the Great Recession in figure 13.[5]

LONG-TERM EFFECTS OF A PAST SHOCK, OR CONTINUING EFFECTS OF A NEVER-ENDING SHOCK? The central motivation of the authors is to assess whether areas most battered by Chinese imports between 2000 and 2010 had been able to adjust by 2019, roughly a decade after Chinese imports reached their plateau. While figure 3, panel B, indeed shows a plateauing of US import penetration from China by 2010, the figure also suggests the possibility of a continuing increase in import exposure when other low-wage Southeast Asian countries to which China may have been offshoring some of its own production over time (such as Vietnam) are included. The fact that the rate of increase in the 2010s is lower than in the 2000s (and somewhat comparable to the 1990s) does not tell us about the distribution of this additional import pressure across CZs. Given this, it would seem relevant for the authors to consider whether the areas that were most exposed to China in the 2000s are also those most exposed to increased import penetration from these low-wage Southeast Asian countries in the 2010s. Absent this, it is more difficult to conclude that the paper is truly looking at the persistence of the (past) China shock rather than the continuing effects of two decades of increased exposure to Southeast Asian import competition.

INTERACTION OF THE CHINA SHOCK WITH THE GREAT RECESSION As already discussed above, the losses of manufacturing jobs in the 2000s were concentrated during the 2000–2002 period and during the Great Recession, while the increase in Chinese imports was continuous during the entire 2000s. The possibility that local labor markets most exposed to the China shock may have had a particularly difficult time adjusting because the shock partly coincided with one of the biggest financial sector–driven recessions in US history deserves more consideration. This seems especially relevant in terms of the generalizability of the findings to other large and sudden

5. The analysis of pre-trend in online appendix figure A1 is limited to the 1991–2000 trade shock, and hence limited to pre-trends up to 1990.

15985-06b_AutorCom&GD-3rdPgs.indd 452 5/12/22 12:06 PM

shocks, such as those one may expect from a fuller energy transition in the United States toward renewables. The US unemployment rate was above 8 percent from 2009 to 2012. Low-skilled workers, who dominate the industrial sectors exposed to Chinese imports, were particularly hurt by the Great Recession, with the unemployment rate among those with less than a high school degree well above 12 percent from 2009 to 2012.

We know from Davis and von Wachter (2011) that the effects of mass layoffs are much more persistent when they occur in a period of high unemployment. Davis and von Wachter (2011) show that losing a job in a mass layoff when the unemployment rate is less than 6 percent results in income losses over the next twenty years that are about one-seventh of counterfactual income. In contrast, losing a job in a mass layoff when the unemployment rate is over 8 percent results in losses twice as large over the next twenty years. In this regard, it is informative that there is a nontrivial correlation between the Great Recession shift-share variable and the China shock shift-share variable (0.48, as mentioned in a footnote). In their quest to understand the sources of the long-term persistence, the authors may have benefited from studying heterogeneity across CZs based on their exposure to the Great Recession.

The Great Recession may have further restricted trade-displaced workers' ability to find new jobs in that, besides constraining the number of job opportunities, it may have also made displaced workers, but especially homeowners, less geographically mobile. The Great Recession was preceded by a sharp increase in homeownership, and many of these homeowners saw the value of their housing wealth decline sharply as the recession hit. Households that would otherwise have found it beneficial to relocate may have been locked in to their existing residence (and town) in that they were not able to sell their home for enough to pay off their lender and would not have had enough cash for a new down payment. Here again, the paper could do more to assess the relevance of this channel to the long-term persistence of the China shock by exploring variation across CZs based on the share of homeowners and their exposure to the housing market collapse.

Putting aside any direct effect of the Great Recession on Americans' ability to relocate, existing research also shows that there has been a substantial downward trend in Americans' mobility over time (Molloy, Smith, and Wozniak 2011). Some of the explanatory factors that have been identified relate to both higher costs of moving (increase in the homeownership rate, aging population, rising share of dual-earn households) but also lower benefits of moving (such as increasing housing costs in job-rich cities and states, or the rise in state- or city-specific occupational licensing). One may

15985-06b_AutorCom&GD-3rdPgs.indd 453 5/12/22 12:06 PM

wonder about additional barriers that are specific to the population and regions that were most exposed to Chinese imports. In many of these disproportionately lower-income communities, extended family and friends may be a crucial part of a private safety net that is difficult to let go of in the absence of a strong government-provided safety net. One may also wonder about the extent of the negative self-selection among those who are still residing in the manufacturing towns that had already experienced decades of decline prior to the China shock.

INDUSTRIAL RELATIONS, CORPORATE GOVERNANCE, AND ACTIVE LABOR MARKET POLICIES Germany's experience with the China trade shock was very different from that of the United States. While the United States witnessed a large trade deficit, Germany maintained balanced trade accounts with China. This was in part because Germany was uniquely able to increase its exports to China while the United States (and many other developed economies) were not. Some of this might be due, as the authors argue, to Germany's unique comparative advantage in a set of specific manufacturing goods for which China's demand was fueled by its economic growth and its rising level of consumer income (such as manufacturing tools and luxury cars). But there might be other useful contrasts between the United States and Germany worth discussing as they raise broader questions for future research. In an interesting piece, Dustmann (2021) speculates as to what might be distinctive about Germany, beyond its industrial structure and much longer history of integration in international trade, especially with other (mostly lower-income) European countries than the United States.

One factor might be Germany's distinctive industrial relations and labor market institutions. One may wonder whether stronger union representation, worker representation on corporate boards, or structures akin to Germany's work councils might have led to different longer-term outcomes and more internalization in business decisions pertaining to the welfare losses for exposed workers, their families, and their community. Relatedly, it would be interesting to know whether firms characterized by local ownership, such as family-held firms, might have been more willing to consider the well-being of their workers when looking for corporate responses to these new competitive pressures.

Also, publicly owned firms and private equity investment are likely more common in the United States than in Germany. These ownership structures might be more focused on short-term profitability and hence less willing to make the investment, including in their workforce, that would be required to reorient future production toward goods for which China does not present

15985-06b_AutorCom&GD-3rdPgs.indd 454 5/12/22 12:06 PM

such a competitive threat, such as higher value-added niche products or products that involve greater customization.

Another distinctive feature of the German economy is its greater preparedness to reskill or upskill its workforce, in particular via its large apprenticeship program. Battisti, Dustmann, and Schönberg (2017) show that German firms that adopt new technological or organizational practices that displace otherwise routine work manage to limit worker losses by retraining or upskilling their workforce. Interestingly, they also show that these positive retraining outcomes are largest for workers in firms that run large apprenticeship training programs and have strong union representation. It would be interesting to know whether the same dynamics of worker reskilling and upskilling happened in Germany in response to the China trade shock.

Finally, and not as much a matter of Germany's than of United States' distinctiveness, government spending on active labor market policies (ALMP)—which include improving job readiness, providing help in finding suitable employment, and expanding employment opportunities—may play an important role in easing labor markets' adjustment process to large and sudden negative shocks. According to the OECD, the United States was ranked second to last among OECD countries in its ALMP spending as a percentage of GDP as of 2017, only ahead of Mexico.[6] The United States spends annually only 0.1 percent of GDP on ALMP, compared to an OECD average of 0.65 percent, and a high of 1.96 percent in Denmark. While evidence on the effectiveness of ALMP is mixed, it is also the case that not much good can happen from spending (close to) nothing to support displaced workers. While such a low level of government spending might have been historically justified by higher levels of internal mobility in the United States, it seems thoroughly inadequate given the decline in American workers' mobility discussed above.

It is true that workers displaced by trade have received disproportionate support, with Trade Adjustment Assistance (TAA) often used as a carrot to win enough support for new trade deals, while no such carrot is needed when purely private decisions, such as the adoption of labor-saving technologies, rather than political decisions are the source of worker displacement. But even here, the level of spending seems inadequate given reasonable expectations about how much it would cost to retrain a worker. The 2002

6. OECD, "Active Labour Market Policies: Connecting People with Jobs," https://www.oecd.org/employment/activation.htm.

TAA program capped training funds at $220 million per year throughout the 2000s when the China shock was full-blown; it was only in 2009 that the cap was raised to $575 million as part of the American Recovery and Reinvestment Act. This amounted to less than $1,700 per new worker certified by TAA petitions in 2007 (Vijaya 2010).

CONCLUSION This paper complements a large literature in labor economics showing that the assumption of perfect labor markets is refuted by the data. This assumption is built into much of government policy beyond trade, such as cost-benefit analysis calculations in rule making or antitrust enforcement. I hope this paper will help contribute to a much needed policy discussion on better accounting for labor market imperfections and worker welfare.

REFERENCES FOR THE BERTRAND COMMENT

Acemoglu, Daron, David H. Autor, David Dorn, Gordon H. Hanson, and Brendan Price. 2016. "Import Competition and the Great US Employment Sag of the 2000s." *Journal of Labor Economics* 34, no. S1: S141–98.

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

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103, no. 6: 2121–68.

Battisti, Michele, Christian Dustmann, and Uta Schönberg. 2017. "Technological and Organizational Change and the Careers of Workers." Working Paper. https://www.qmul.ac.uk/sef/media/econ/events/Technological-and-Organizational-Change-and-the-Careers-of-Workers.pdf.

Binder, Ariel J., and John Bound. 2019. "The Declining Labor Market Prospects of Less-Educated Men." *Journal of Economic Perspectives* 33, no. 2: 163–90.

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

Dustmann, Christian. 2021. "Trade, Labor Markets, and the China Shock: What Can Be Learned from the German Experience?" Discussion Paper 2112. London: Centre for Research and Analysis of Migration.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *Journal of Economic Perspectives* 25, no. 3: 173–96.

Pierce, Justin R., and Peter K. Schott. 2016. "The Surprisingly Swift Decline of US Manufacturing Employment." *American Economic Review* 106, no. 7: 1632–62.

Vijaya, Ramya M. 2010. *Broken Buffer: How Trade Adjustment Assistance Fails American Workers*. New York: Demos.

Wang, Zhi, Shang-Jin Wei, Xinding Yu, and Kunfu Zhu. 2018. "Re-examining the Effects of Trading with China on Local Labor Markets: A Supply Chain Perspective." Working Paper 24886. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w24886.

15985-06b_AutorCom&GD-3rdPgs.indd 456 5/12/22 12:06 PM

COMMENT BY

EDWARD L. GLAESER The China shock literature, pioneered by Autor, Dorn, and Hanson (2016), now represents a major strand within two fields of economic thought: international trade and regional economics. As this new paper shows, after twenty years there has been virtually no recovery in employment, manufacturing employment, total earnings, or per capita earnings of the regions experiencing the most impact from the shock. The economists of the 1990s and early 2000s, which sadly includes myself, who were almost unalloyed boosters of free trade rarely imagined that either NAFTA or the admission of China into the World Trade Organization (WTO) could lead to such enduring regional dysfunction. The great puzzle for regional economists posed by the facts documented in this paper, and the broader China shock literature, is why did large parts of America prove so unable to adapt. Why was there so little regional resilience?

In this comment, I highlight four hypotheses about America's regional sclerosis: (1) the shock just keeps going and going, like the Energizer Bunny; (2) the shock struck areas that lacked protective human capital, which can be formal schooling, entrepreneurship, or a diversity of industries; (3) America has acquired local institutions that reduce the ability of repurposing labor and capital; and (4) migration no longer enables areas to rightsize themselves in response to negative shocks. As it stands, the paper superbly documents the persistence of the China shock and provides suggestive evidence on the duration of the shock and the absence of protective human capital. Yet there is a clear need for future research to go forward and help us to better understand why the China shock was so persistent.

The stakes are indeed very high, both for current public policy and for planning for future trade events. Understanding the source of America's regional malaise does not immediately imply a policy solution, but it helps point to potential remedies. For example, if the primary problem was that this shock was extremely persistent, then perhaps it would make sense to focus primarily on short-term relief, as we do now, trusting that eventually the shock would end. If the problem lies in regional human capital, then strategies to improve education and entrepreneurship might be warranted. If the problem lies in regulatory limits on adaption, then this suggests removing the barriers to business formation. If the problem lies in migration, then it may make sense to focus on the barriers to building in America's more productive regions.

The value of resolving this puzzle for future trade policy is even more obvious. Many of us were blindsided by the size of the China shock. If we understood better what makes a region susceptible to trade competition,

then we can predict whether a future shock will be more or less disruptive. We would then be able to make better decisions about when and whether to open, and what policies are needed to deal with the downsides of having new trading partners.

I begin by discussing why the lack of local resilience to the China shock is such a puzzle. I then turn to the four hypotheses, explaining them in more detail and describing the empirical implications of each, some of which are already in the paper. I then discuss a possible route toward a decomposition.

WHY IS THE PERSISTENCE OF THE CHINA SHOCK A PUZZLE? Before 1992, economists who specialize in both international trade and regional economics rarely expected that whole regions of America would lurch into crisis if the United States opened itself more to Mexico (NAFTA in 1994) or China (accession to the WTO in 2001). Trade economists certainly knew that America's low-wage workers could lose out from competition with poorer countries, but they rarely considered the regional implications of that fact. Moreover, the vast literature on the rise in returns to skill was just starting, and there was no consensus (and really still isn't) on the role that globalization played in pushing down the wages of less well-educated Americans. Since Adam Smith, the majority of economists have typically disliked economic nationalism, championed free trade, and consequently downplayed trade's potential downsides in our more popular writing.

As I am a regional economist and no expert on international trade, I will focus on why the persistence of the shock was so surprising to those of us who study local economies. Most obviously, America is a large country, and its tariff walls have been high throughout its history. Consequently, trade has played an outsized role in the success of individual cities, which specialized in shipping and commerce, but negative trade shocks rarely if ever did damage to larger regions of the economy prior to 1970.

Major regional shocks came from transportation innovations, like canals and railroads, new technologies, like the cotton gin and the air conditioner, and political changes, like the right-to-work laws that attracted manufacturing after World War II (Holmes 1997). After 1970, it was obvious that competition with Japan was causing difficulty for some manufacturing sectors, such as automobile production. Yet regional economists were more likely to note continuity than change and to see the decline of industrial cities, like Detroit, as a long-running process reflecting automation and the post–World War II relocation of factories to lower-wage areas.

Moreover, regional economists were largely persuaded by Barro and Sala-i-Martin (1992) that income convergence at the place level was

something of a universal law, true across countries and regions and true for over a century. Blanchard and Katz (1992) also highlighted that unemployment rates do not persist across states, partially because healthy emigration from depressed areas led wages in those areas to rise. Even though Decressin (1994) found that in Europe the process of regional adjustment occurred through the much less healthy process of prime-age men leaving the labor force altogether and Glaeser and Maré (2001) documented strong thirty-year persistence of unemployment rates at the city level, these findings seemed irrelevant for larger American regions, which were dynamic and ever-changing.

Moreover, the prevailing theories of regional economics emphasized the existence of a spatial equilibrium in which welfare levels for equivalent individuals were equalized across space. Alonso (1964) first applied this idea to the land market within the city. Rosen (1979) and Roback (1982) applied the idea formally across metropolitan areas, although the basic idea is also present in Harris and Todaro (1970). The power of movement to iron out differences of welfare across space seemed to be guaranteed by the more than 6 percent of Americans who moved across county lines during every year from 1950 to 1992 (Molloy, Smith, and Wozniak 2011).

Spatial equilibrium implies that even if less skilled workers in America as a whole are worse off because of the change in the terms of trade, these workers should not be disproportionately harmed just because they live in one region or another. Going forward, it would be extremely useful if the China shock literature could document that the residents of shocked areas suffer even more than their level of education or industrial expertise would suggest. Linked census and tax records would be ideal for answering this question.

The Harris-Todaro model explains persistent unemployment at the place level with a high minimum wage. That wage attracts migration to the point that expected income is equalized to earnings in the low-wage region. As wages are fixed in one region above wages in the other region, the probability of employment in the high-wage region must be less than one. Hall (1974) took this model to the US data and showed that high earnings and high unemployment rates went together across the United States, perhaps because midwestern unions were so effective in keeping wages up during the early 1970s.

Gradually models were developed with barriers to migration, such as the place-specific taste shocks that have become ubiquitous in the urban literature since Diamond (2016). Other models emphasized the impact of place on human capital formation, which could induce patterns of poverty across

generations (Benabou 1993; Glaeser 1998). But these papers focused more on small neighborhoods than on whole regions.

The nonmathematical and historical tradition in regional analysis similarly emphasized dynamism rather than stasis. While studies of European regions would often seek explanations for perpetual backwardness, for example, Sicily (Banfield 1958), American regions seem to adapt to shocks. Our Mezzogiorno was the Deep South, and by the 1960s it seemed to change rapidly both culturally and economically. Smith and Welch (1989) showed that the Great Migration of African Americans out of the South explained much of the postwar income convergence between African Americans and white people.

One final fact also encouraged the belief that shocked local economies regularly reinvent themselves: the recovery of some (but hardly all) cities from the urban crisis of the 1970s (Glaeser 2011). Slammed by deindustrialization, suburbanization, and rising crime levels, cities from New York to Seattle seemed like they were facing an existential threat in the 1970s. By the early 1990s, it was obvious that the better-educated citizens of these cities had reinvented them from cities that made goods to cities that specialized in finance and technology. As these positive stories seemed far more interesting than the persistent poverty of less well-educated cities, like Cleveland, the narrative of their recovery played an outsized role in the thinking of many experts on local economies. Even the fate of Detroit seemed like evidence for the spatial equilibrium model, for while its wages did not recover, its housing costs were extraordinarily low. Detroit provided evidence for the durable housing model of Glaeser and Gyourko (2005), which predicts that urban decline can go on for decades because durable housing remains even after the jobs disappear. But that kind of permanent decline seemed far less plausible for whole regions of the United States.

FOUR HYPOTHESES ABOUT THE PERSISTENCE OF THE CHINA SHOCK The following four hypotheses about why the China shock has been so persistent at the place level seem like the most natural explanations, but they are not the only hypotheses.

The China shock just keeps going and going. The most natural way to think about a trade event is essentially as a break in time. Before 2001, the global trading network was in equilibrium without China. After 2001, the global trading network was in equilibrium with China. If everything about the world is static other than that break, then this represents a single shock, and we would expect the world to get to a new equilibrium reasonably quickly. From this perspective, the enduring impact of the China shock on

many American regions is a puzzle. Why aren't they coming to some new equilibrium?

But of course, the China shock does not seem to have been a one-off event. China's entry into the global trading system was followed by massive economic growth accompanied by massive technological development and improvements in education. Consequently, China started with a comparative advantage in goods made with inexpensive labor but then gradually increased its productive edge in a wider range of industries. The shock then was more like rolling thunder than a single lightning strike.

At the regional level, this changing threat may have moved from one industry into another. Firms displaced from a first industry could have moved to a second, which then experienced the full brunt of Chinese competition two years later. Consequently, there may have been considerable adjustment to the China shock, but that adjustment couldn't heal the region because each adaptation was hit by China expanding into a new sector.

The authors do implicitly address this hypothesis at the start of the paper, even if it is not labeled. Figure 1 shows that China's share of world trade continues rising through 2015, which would imply that American regions needed to keep adapting to changing conditions for fifteen years and have had only five years to react to the new equilibrium. If that is the right view of the China shock, then it would seem like this hypothesis is largely correct.

However, the authors embrace Balassa's (1965) measure of revealed comparative advantage for manufacturing, which is a country's share of global manufacturing exports relative to the country's share of world exports of all other goods. As China's total share of global manufacturing exports actually declined between 2010 and 2018, China's revealed comparative advantage in manufacturing relative to the United States stops rising steeply after 2009 as well. Consequently, they see 2010 as the end of China's "productivity growth miracle," which would mean that America has now had over a decade to react, which would make the persistent problems of impacted regions more puzzling.

At this point, I remained unconvinced as to whether the turning point is 2015, as suggested by China's overall share of global exports, or 2010, as suggested by China's revealed comparative advantage in manufacturing. As I am not an international trade economist, I admire Eaton and Kortum's (2002) model immensely, but I do not take it as an article of faith (unlike certain sections of Adam Smith's *The Wealth of Nations*). The figures shown in the paper do not rule out the possibility that China's manufacturing

powerhouse was continuing to expand through 2015, although its nonmanufacturing exports were as well.

The authors should address this far more convincingly. Setser (2020) shows that China has continued to rack up ever larger balance of payment surpluses (with the world, not particularly with the United States) through 2015, which is certainly compatible with the shock continuing through that year. The balance of payments surplus is rising, even as exports are stagnant, because Chinese manufacturing imports are declining. This is compatible with a view that China learned how to make a wider range of products between 2010 and 2015 and consequently no longer needed to import them. That could easily have led to more shocks to the already troubled regions in the United States.

I hope that the authors can settle this more decisively going forward, and in the interest of a more exciting puzzle, I hope that they are right that the shock ended earlier. I suggest here a few relatively easy steps to address this topic, but I am sure that they can come up with other, better ways of showing when the China shock came to an end. Most obviously, they can present direct data on annual productivity growth in the manufacturing sector within China and the United States.

A second alternative is to focus on a set of industries within the United States that experienced the greatest impact from the China shock. Was the decline in employment in those industries largely over by 2010, or did it continue for another five years? What happened to total factor productivity in those industries both in the United States and in China over this time period?

A third possibility is to focus on the range of industries in which China expanded. How many different types of goods were being exported successfully by China over time? A simple breakdown of Chinese employment by industrial subclass should work here, with again a focus on the 2007–2016 period that is the subject of contention. Presumably, these facts relate to the decline in China's manufacturing imports, so it might be reasonable to show also the time series of those facts over the same period.

A fourth area is China's nonmanufacturing exports. In a purely static model, of course, China cannot be gaining ground relative to the United States in both manufacturing and nonmanufacturing sectors. In a model with substantial short-term trade deficits, it is quite possible for China to be displacing American exports across the board. Consequently, it would be quite useful to see some data on the time series of nonmanufacturing exports from both countries and whether it is possible that the surge in China's nonmanufacturing exports after 2010 could have also contributed to the malaise of America's underperforming regions.

15985-06b_AutorCom&GD-3rdPgs.indd 462 5/12/22 12:06 PM

I focused here on approaches that should be straightforward to implement and collectively more convincing about the time when the China shock came to an end. Others may have much better ideas about how to make this point more clearly. Yet I think it is vital to do a better job of determining whether the China shock ended in 2010 or before or in 2015 or after.

The dearth of human capital hypothesis. Eriksson and others (2021) takes a product cycle approach to regions in which industries move from innovative, high-education regions to low-wage, low-innovation regions and found that "the correlations by 1990 suggest that the China shock industries were no longer leading industries concentrated in high-education areas, but had moved through the product cycle" (6). A hypothesis, associated above all with T. W. Schultz and Finis Welch, is that education or other adaptable skills enable people and regions to adapt to changing conditions. The fact that the China shock hit regions with less-educated labor forces may have meant that these regions lacked the skills needed to reinvent themselves. Eriksson and others (2021) do indeed find that the China shock had far more severe consequences on labor force participation in areas with a less-educated workforce; and by contrast, they found that "the Japan shock was on average hitting communities that had a population with a higher level of education and innovative capacity" and consequently "this may have given those communities a better opportunity to pivot to other industries and tasks" (11).

The regional and urban literature has interpreted local human capital broadly to include formal education, entrepreneurial human capital, and a diversity of industrial skills. For example, Glaeser and Saiz (2004) examine the idea that education mediates the negative effect of cold weather on urban population growth from 1940 to 2000 and find a strong negative interaction between years of schooling and warmth, meaning that education appeared to enable reinvention. Similarly, they found that education in 1940 and the interaction between education and manufacturing in 1940 both predicted the decline in manufacturing between 1940 and 2000, suggesting that education enabled cities to find postindustrial areas of specialization.

Chinitz (1961) emphasized entrepreneurial, rather than formal, human capital in explaining why New York seemed to be more resilient than Pittsburgh. New York had been dominated by the garment industries, which had weak returns to scale and few barriers to entry. Pittsburgh because of its proximity to coal mines, specialized in steel, which at its heyday was practically a natural monopoly. If entrepreneurial human capital is particularly fungible across industries, then it will enable places to develop new lines of business after an event like the China shock. Typically, entrepreneurship

is measured with either average establishment size or share of employment in new establishments at the beginning of the time period, and it is sometimes instrumented for with natural variables, such as proximity to historical mines.

Industrial diversity is also seen as a potential form of local human capital, which both engenders creativity and enables reinvention. Jacobs (1969) emphasized cross-industry leaps of imagination, which she claimed were particularly prevalent in diverse cities like New York, and Weitzman (1998) formalized that idea. The literature on local diversity is less well developed than the work on years of schooling or entrepreneurship, but papers do find that initial industrial diversity does predict subsequent employment and population growth at the local level.

The authors have already taken the first steps to test this hypothesis in figures 9 and 10, which seem to show support for the importance of local human capital. Figure 9 focuses on the interaction between the China shock and formal years of education. The first panel looks at the ratio of manufacturing employment to working-age population, which seems to decline equivalently in both educated and less educated regions. By contrast, the ratio of employment to population drops substantially in the less educated regions and actually increases in more educated regions. This may be the single most striking and surprising fact within the paper, and it strongly suggests that local skills shaped the impact of the China shock. The third panel shows that the China shock led to a decline in population in the regions with more educated populations, but not in the less educated areas. The fourth panel shows that the China shock had a milder impact on income in less educated areas, but that the impact persisted and if anything became more severe over time. In educated areas, the initial income drop was severe, but it had essentially disappeared by the end of the time period.

In short, figure 9 suggests that the dysfunction occurred almost entirely in less educated parts of America. Educated workers reacted by leaving the region. That out-migration combined with whatever other adaptations were going on led the initial terrible shock to income to right itself. By contrast, the response in less educated areas was to leave the labor force entirely. The growth of American joblessness is one of the most terrible correlates of the China shock, although that trend preceded the shock by thirty years.

These findings can be expanded and made slightly more formal (i.e., more *t*-tests between the two groups or systematic interactions in regressions). Again, it would be good to know if this was about the people or the place,

which requires some form of correction for individual attributes. Finally, it would be nice to know about other forms of adaption such as changes in the industrial composition at the local level.

Nonetheless, these findings at least suggest that the China shock may have been so persistent because it disproportionately struck less-skilled places. In these areas, the less well educated were closer to the margin of not working and the shock pushed them over the edge. That caused earnings to fall further because of reduced demand for local services, and more people may have lost their job. There was also less human capital that would enable new business formation. In this view, the surprising persistence of the China shock represents the fact that it struck the most vulnerable parts of America. Those regions then developed a form of the unemployment hysteresis that was first observed in Europe in the 1980s and 1990s. America's formal and informal safety nets (i.e., friends and family) are not generous enough to keep the skilled unemployed from trying to get another job, but maybe they have become generous enough so that the unskilled who are laid off just leave the labor force altogether.

The results on specialization shown in figure 10 suggest that the shock was worse for more-specialized metropolitan areas. Yet it is not clear if this is because these geographic areas were smaller, or if the same size shock was effectively more severe in the more specialized areas. It is also difficult to understand the interaction between a shock based on industrial composition and another measure of industrial composition. As much as I appreciate the transparent figures produced by the authors, this is a case where it would make sense to do this in a regression framework where we can tell if an apparent interaction with specialization is actually an interaction with the size of the agglomeration. Presumably it would also make sense to know if there is an interaction with specialization controlling for the interaction with education.

If the concentration interaction remains robust, then a few further steps seem warranted. First, it is worthwhile distinguishing between concentration in industries that were particularly struck by the China shock versus industries that were not. The level of concentration in China shock industries might well have determined the size of the initial blow. The concentration of alternative industries would presumably influence the ability to branch into multiple different new industries.

Finally, just as it makes sense to look at whether education is helping produce more of an industrial shift in response to the China shock, it makes sense to see whether diversity is allowing more of a reaction. Again, in this case, the natural outcome is some measure of new growth in export

industries—exportable from the commuting zone (CZ), not from the United States—that are not shocked by China. Naturally, it would also make sense to look at horse race regressions with this alternative measure.

The last measure of local human capital is entrepreneurship. In this case, the authors could just take share of employment in new firms or average establishment size or both in 2000 or earlier and again look at interactions between these variables and the China shock. Presumably the same steps should be taken for all three concepts of local human capital, which should be relatively straightforward.

The barriers to adjustment hypothesis. The third hypothesis is straightforward to describe but difficult to test. March and Olsen (1998) argued that small well-organized interest groups gradually manage to acquire power in peaceful societies. Consequently, societies accrete rules that protect those insiders at the expense of outsiders, and these rules limit growth in the country as a whole. He saw Britain in the 1970s as a prime example of a nation where insiders, including the clubby financiers in London and union members, had locked out newcomers and held down the country's performance. Of course, the big bang that deregulated London's stock exchange in 1986 seems to prove both that these rules need not be an absorbing state and that Britain's pre-1986 financial rules were limiting the growth of its financial sector.

In the United States, occupational licensing has expanded dramatically since the 1950s and that can make it more difficult for people to switch occupations or to move across state lines within the same occupation (Kleiner and Krueger 2010). In many cities, a vast number of regulations make it difficult to open new businesses that sell physical goods. Indeed, America seems to regulate the entrepreneurship of the poor, which largely involves physical products, far more tightly than it regulates the entrepreneurship of the rich, which often occurs in cyberspace. The barriers to new home construction can also be seen as rules that protect insiders at the expense of outsiders, but they are part of the next hypothesis.

While I believe that this remains a valid hypothesis, there are only small steps that can be taken to address it at this point. The simplest approach, which is also less than satisfying, is to use the data that Morris Kleiner has assembled on occupational licensing at the state level and to see whether there is any interaction with the China shock. If there is no significant effect, which I suspect will be the answer, then hopefully future work can provide a possible explanation. A similar step could be taken with the various indexes of the regulation of small businesses at the state and local level (Winegarden 2015).

15985-06b_AutorCom&GD-3rdPgs.indd 466 5/12/22 12:06 PM

A far more compelling approach would be to look at individual panel data and movement across states and occupations as a function of the presence of licensing, but that is far beyond the scope of this paper. This would require better data on the state of occupational licensing than I believe that we possess, as well as a data set that enables us to look at the mobility of large numbers of individuals across occupations. Nonetheless, the basic question would be whether the need to license deters people who are hit with the China shock from moving into particular occupations.

Ideally, we could form an overall level of licensing based on the occupations that attract workers who are affected by the China shock. We could ask whether the China shock has a larger impact on joblessness in states where this measure of licensing severity is more extreme. We could also form different indexes for men and women, or other specific groups, and then get more variation in the interaction between exposure to the China shock and exposure to a more stringent licensing regime.

A similar path could be taken with new business regulation. If there was a measure of stringency of regulation by industry and location, then we could form a state or locality level of exposure to this regulation based on the industries that typically attract people who are affected by the China shock. That measure could again be interacted with the China shock, and we could test whether the China shock did more damage in places that regulate small business growth more severely.

The closing of the metropolitan frontier. The final hypothesis is an alternative interpretation of the previous one: insiders have made it impossible to afford housing in high-wage metropolitan areas, which means that displaced workers cannot move to find opportunity. Ganong and Shoag (2017) show that directed migration of poor people from low-wage states to high-wage states largely disappeared after 1980. One plausible reason for this is that high-wage states make it difficult to build low-cost housing, and consequently the costs of living eliminate any wage gains that would come from mobility.

This fact similarly helps explain the rise in the returns for low-skilled workers found in Autor (2019). In fact, the return to urban agglomeration remained as strong in 2010 as it was in 1970 for low-skilled workers who live in high-skill metropolitan areas. However, the returns to living in large, low-skill metropolitan areas have completely vanished, and low-skilled individuals naturally live disproportionately in low-skill and low-wage metropolitan areas. Those metropolitan areas have low living costs, largely because durable housing remains long after an area's productivity vanishes (Glaeser and Gyourko 2005). Conversely, well-educated people are particularly good

at erecting barriers to new building, and consequently the limits on construction are particularly fierce in highly educated metropolitan areas and the prices of homes are also particularly high.

As figure 9 shows, population levels decline in more-skilled CZs and then per capita personal incomes increase. That recovery could reflect a tighter labor market for the remaining workers or selective out-migration by those who were most severely affected by the China shock. Yet there is no population loss from the less-skilled areas that were hit by the China shock, and that may reflect the limited supply of housing in attractive migration sites.

How could we test this hypothesis? The rise in housing regulation is nationwide, but there are stronger migration linkages between certain metropolitan areas. Schubert's (2021) job market paper studies the flow of housing market contagion across sites that are linked by migration networks. As these migration networks have been measured, it is relatively straightforward to measure the migration-weighted housing price index for every CZ. At a point in time, that index would be higher if a particular city is more likely to send people to areas that have higher prices. The index would be easy to adjust over time, and so consequently it would be possible to test whether population out-migration from place A is higher when the average price index of its linked areas is lower. It would also be easy to check if the China shock's impact on population was larger during periods when a place's migration price index was lower.

A somewhat better measure would take into account both the prices and the wages in the migration-linked area. The cost of living could be subtracted from the average earnings of less-skilled households. That total measure could then be weighted by migration links for each CZ. Again, the goal would be to see if there was more population loss when a CZ's migration links offered a better bundle of earnings minus costs.

Alternatively, the American Community Survey could be used to look at actual migration patterns. These could be linked with housing costs and the China shock. An individual migration regression could then be run that tests whether the China shock induces significant migration to high-wage, high-cost areas. The absence of such an effect would also provide evidence for the hypothesis that limited housing supply is preventing people from moving out of depressed areas.

TWO DECOMPOSITIONS OF THE CHINA SHOCK'S PERSISTENCE As the persistence of the China shock is unlikely to reflect any one of these hypotheses alone, it would be nice to have some way of apportioning credit to the

various causes. Two different decompositions might help allocate the quantitative importance of the different hypotheses.

Decomposition 1: Timing, geography, and out-migration. The first decomposition begins with a simple, macro-like equation linking the economic outcome of interest, such as the employment to population ratio, with the China shock and other attributes:

$$y_t^{cz} = \sum_j \left(\beta_{0,j}^{CS} + \sum_a \beta_{a,j}^{CS} X_a^{cz}\right) CS_{t-j}^{cz} + Other\ Controls + \varepsilon_t^{cz}.$$

In this equation, y_t^{cz} is the outcome; $\beta_{0,j}^{CS}$ is the direct effect of the China shock occurring j periods ago; $\beta_{a,j}^{CS}$ are various interactions with local attributes such as education, entrepreneurship, antibusiness regulations, and industrial diversity; X_a^{cz} are the time-invariant levels of the attributes; and ε_t^{cz} is the error term.

In order to take a stand on the first hypothesis some view about how shocks decay over time is needed. One option is to assume that $\beta_{i,j}^{CS} = \delta^j \beta_{i,0}^{CS}$ for all of the attributes and then to use an estimate of δ either pulled from the literature or taken from parallel studies of coal and the business cycle.

The first natural decomposition is to ask how different would America look today if the China shock had been shorter lived or if the China shock had hit different parts of America. How much lower would regional equality be if, instead of hitting the places that had these industries in 2000, the China shock had hit the places that had these industries in 1970 or 1910—at least according to Eriksson and others (2021)? These thought experiments seem straightforward, and if the interactions are to be believed, they would be pretty interesting.

The lightest version of this occurs if only one interaction (education) is considered and timing is not a concern. In this case, one would simply take the basic China shock results and ask how different America would be today if the China shock had struck different geographies. This is close to the authors' comparison figures between more or less educated areas, but it gives us a chance to be somewhat more systematic.

To add in the effects of the closing of the metropolitan frontier, we would need to take a stand on the impact of out-migration on the outcome of interest. The equation might instead take the form:

$$y_t^{cz} = \sum_j \left(\beta_{0,j}^{CS} + \sum_a \beta_{a,j}^{CS} X_a^{cz}\right) CS_{t-j}^{cz} + \theta ln\left(\frac{Pop_t^{cz}}{Pop_{2000}^{cz}}\right) + Other\ Controls + \varepsilon_t^{cz}.$$

In this formulation, θ would presumably be negative for out-migration to be doing any good to the local economy. The right step then is to simulate how many people would have left the CZ if there was more building in the high-value metropolitan areas. Yet to do that properly, a full economic geography model would be needed, perhaps along the lines of Diamond (2016). Consequently, this is surely best left for future researchers.

Decomposition 2. The second decomposition that might be useful is focused just on employment flows, not on the four hypotheses. The goal of this decomposition is to look at the total change in employment in the impacted sectors and then apportion that change to out-migration, employment growth in other sectors (perhaps by broad category), and nonemployment. This would only really make sense within a particular age group, either 18–64 (which the authors use) or the standard prime-age group, which is 25–54.

Using the equality $Pop_t^{cz} = MEmp_t^{cz} + OEmp_t^{cz} + NEmp_t^{cz}$, where $MEmp_t^{cz}$, $OEmp_t^{cz}$, and $NEmp_t^{cz}$ stand for manufacturing employment, other employment, and nonemployment, we get:

$$\frac{MEmp_{2000}^{cz} - MEmp_{2020}^{cz}}{Pop_{2000}^{cz}} = \frac{Pop_{2000}^{cz} - Pop_{2020}^{cz}}{Pop_{2000}^{cz}} + \frac{OEmp_{2020}^{cz} - OEmp_{2000}^{cz}}{Pop_{2000}^{cz}} + \frac{NEmp_{2020}^{cz} - NEmp_{2000}^{cz}}{Pop_{2000}^{cz}}.$$

These three categories can then be related both to the China shock and the various interactions. This is not far from what the authors have done, but it helps us to see clearly how much of the manufacturing change is related to nonemployment, out-migration, and the growth of other industries.

At the least, I would again separate out other employment into local service employment (which should certainly decline) and other export industries. This decomposition is useful if readers want to think about whether the local attributes are influencing the growth of employment in other sectors.

CONCLUDING THOUGHTS The work of Autor, Dorn, and Hanson on the China shock is of first-order importance both to economic science and public policy. It matters for trade policy and it matters for policies that relate to the differing fortunes of America's regions. We are at a point where regional heterogeneity feels particularly painful, especially in the political realm, and so this work is incredibly useful.

Looking forward, I think it is important that future regional work on the China shock differentiate three broad classes of outcomes: (1) those outcomes that directly relate to welfare (such as income and nonemployment rates), (2) those outcomes that capture the direct impact of the China shock itself but do not have any immediate welfare interpretation (employment in the most impacted industries or employment in manufacturing), and (3) attendant outcomes that help to test hypotheses about why the China shock persists or why it has such significant impacts on welfare. All of these outcomes are interesting, but not all of them are equally important. Some form of model is needed to produce coherent estimates of welfare losses or total impact.

For example, total area population fits in the third category unless you want to take a stand on a spatial model where population loss shows that the region as a whole has become less attractive. In that case it should be always joined together with housing prices. Income, population, and housing prices are typically used together in regional work that attempts to sort out whether local changes are coming from declining productivity, changing local amenities, or more abundant labor supply. Traditional regional work assumed perfect factor mobility for labor and capital. Social welfare was captured by the returns to and value of the geographically fixed factors of production and consumption, such as land.

More sophisticated modern work on economic geography has assumed limited mobility of labor and idiosyncratic tastes for particular locations. With those assumptions, labor becomes quasi-fixed, and consequently workers can suffer when their region goes into economic distress. Hopefully, future work in this area will use the new tools of spatial economics and structural estimation to better understand the welfare consequences of place-based shocks.

REFERENCES FOR THE GLAESER COMMENT

Alonso, William. 1964. "The Historic and the Structural Theories of Urban Form: Their Implications for Urban Renewal." *Land Economics* 40, no. 2: 227–31.

Autor, David H. 2019. "Work of the Past, Work of the Future." Working Paper 25588. Cambridge, Mass.: National Bureau of Economic Research. https://www.nber.org/papers/w25588.

Autor, David H., David Dorn, and Gordon H. Hanson. 2016. "The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade." *Annual Review of Economics* 8, no. 1: 205–40.

Balassa, Bela. 1965. "Trade Liberalisation and 'Revealed' Comparative Advantage." *Manchester School* 33, no. 2: 99–123.

Banfield, Edward C. 1958. *The Moral Basis of a Backward Society*. New York: Free Press.

Barro, Robert J., and Xavier Sala-i-Martin. 1992. "Convergence." *Journal of Political Economy* 100, no. 2: 223–51.
Benabou, Roland. 1993. "Workings of a City: Location, Education, and Production." *Quarterly Journal of Economics* 108, no. 3: 619–52.
Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity* no. 1: 1–75.
Chinitz, Benjamin. 1961. "Contrasts in Agglomeration: New York and Pittsburgh." *American Economic Review* 51, no. 2: 279–89.
Decressin, Jörg W. 1994. "Internal Migration in West Germany and Implications for East-West Salary Convergence." *Weltwirtschaftliches Archiv* [*Review of World Economics*] 130, no. 2: 231–57.
Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000." *American Economic Review* 106, no. 3: 479–524.
Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70, no. 5: 1741–79.
Eriksson, Katherine, Katheryn Russ, Jay C. Shambaugh, and Minfei Xu. 2021. "Reprint: Trade Shocks and the Shifting Landscape of US Manufacturing." *Journal of International Money and Finance* 114:102407.
Ganong, Peter, and Daniel Shoag. 2017. "Why Has Regional Income Convergence in the US Declined?" *Journal of Urban Economics* 102:76–90.
Glaeser, Edward L. 1998. "Are Cities Dying?" *Journal of Economic Perspectives* 12, no. 2: 139–60.
Glaeser, Edward L. 2011. *Triumph of the City*. New York: Penguin Books.
Glaeser, Edward L., and Joseph Gyourko. 2005. "Urban Decline and Durable Housing." *Journal of Political Economy* 113, no. 2: 345–75.
Glaeser, Edward L., and David C. Maré. 2001. "Cities and Skills." *Journal of Labor Economics* 19, no. 2: 316–42.
Glaeser, Edward L., and Albert Saiz. 2004. "The Rise of the Skilled City." *Brookings-Wharton Papers on Urban Affairs* 5: 47–105.
Hall, Robert E. 1974. "The Process of Inflation in the Labor Market." *Brookings Papers on Economic Activity*, no. 2: 343–93.
Harris, John R., and Michael P. Todaro. 1970. "Migration, Unemployment and Development: A Two-Sector Analysis." *American Economic Review* 60, no. 1: 126–42.
Holmes, Thomas J. 1997. "Industrial Policy and Competing Jurisdictions." Paper prepared for the Conference on the Law and Economics of Federalism, University of Minnesota Law School and Federal Reserve Bank of Minneapolis, May 2–3.
Jacobs, Jane. 1969. *The Economy of Cities*. New York: Random House.
Kleiner, Morris M., and Alan B. Krueger. 2010. "The Prevalence and Effects of Occupational Licensing." *British Journal of Industrial Relations* 48, no. 4: 676–87.
March, James G., and Johan P. Olsen. 1998. "The Institutional Dynamics of International Political Orders." *International Organization* 52, no. 4: 943–69.
Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *Journal of Economic Perspectives* 25, no. 3: 173–96.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90, no. 6: 1257–78.
Rosen, Harvey S. 1979. "Housing Decisions and the U.S. Income Tax: An Econometric Analysis." *Journal of Public Economics* 11, no. 1: 1–23.
Schubert, Gregor. 2021. "House Price Contagion and US City Migration Networks." Job Market Paper. Cambridge, Mass.: Joint Center for Housing Studies, Harvard University.
Setser, Brad W. 2020. "Looking Back at China's 2019 Balance of Payments Data." New York: Council on Foreign Relations.
Smith, James P., and Finis R. Welch. 1989. "Black Economic Progress after Myrdal." *Journal of Economic Literature* 27, no. 2: 519–64.
Weitzman, Martin L. 1998. "Recombinant Growth." *Quarterly Journal of Economics* 113, no. 2: 331–60.
Winegarden, Wayne. 2015. *The 50-State Small Business Regulation Index*. San Francisco: Pacific Research Institute.

GENERAL DISCUSSION Austan Goolsbee first offered a summary of the paper, noting that the persistent shocks documented in the paper are tied to communities that were highly concentrated in lower-skill manufacturing industries and that the impact that comes from the China shock specifically needs to be separated from the long-term trend of decline in manufacturing employment—was a decline inevitable and merely happened sooner from the shock? Goolsbee also wondered what it was that seemed to make adjustment in US low-skill manufacturing locations so much worse than elsewhere in the world. Goolsbee noted that Chinese competition was predominantly in lower-skill manufactured goods; but the US manufacturing share of employment is actually quite low, and lower-skilled manufacturing is only a fraction of that, whereas in a country like Mexico, manufacturing is a large share of employment and lower-skill manufacturing is the largest component. Goolsbee argued that with so much overlapping of what China exports and what Mexico manufactures, it would seem like the China shock should have been substantially larger on Mexico than on the United States. But Goolsbee then referred to a paper by economists at the Inter-American Development Bank that suggests only a modest impact of the China shock on Mexican labor markets.[1] He said he could not understand why US labor markets would be less flexible than in Mexico following a smaller shock.

1. Juan Blyde, Matias Busso, Valeria Faggioni, and Dario Romero, "The Impact of Chinese Competition on Mexican Labor Outcomes," working paper, 2017, https://www2.gwu.edu/~iiep/waits/documents/Blyde.pdf.

Heather Boushey reiterated Goolsbee's prior question on factors contributing to the varying impacts across countries. She also found it interesting that the authors applied this method to the coal community and questioned whether the authors had considered the implications for other sectors that might be affected by the transition to zero-carbon emissions, for instance, automobiles, other manufacturing, cement, or steel. She highlighted that the list of sectors that will be affected by climate change is extensive, so the authors' work may have broader application to those reallocation effects.

Gordon Hanson first thanked both discussants, Marianne Bertrand and Edward Glaeser, for their comments. Hanson then discussed the German example, which Bertrand had highlighted, where differences exist in labor market institutions between the United States and Germany. He also agreed that considering a longer time horizon, accounting for both regulatory and entrepreneurial factors, as well as labor market institutions, is key to understanding resiliency differences across countries.

Hanson then expanded on two other significant differences between the United States and Germany over this time period. First, the particular role of Germany in the European Union and euro area allowed it to sustain structural trade surpluses, while the opposite was true of the United States, which instead runs trade deficits partially due to the reserve feature of the dollar.[2] Those differences bias the nature of adjustment to be in favor of manufacturing in Germany, while they bias against manufacturing in the US context. The second point he raised considers global supply chains. In Germany supply chains run through manufacturing, where the export side of the German economy that is benefiting from trade with China is largely physical goods, like machine tools.[3] In contrast, US trade with China benefits the technology services sector. Hanson concluded that this is the reason that adding exports of manufacturing to China into the model results in nearly no change. He then highlights that exports of technology services are harder to measure and often end up locked in patent boxes in Ireland and other places. Exports of technology services still confer substantial benefits to regions in the United States and other developed countries that are highly concentrated in those industries; however, these

2. David Dorn, "Commentary: The Impact of Trade on Inequality in Developing Countries." In *Economic Policy Symposium Proceedings: Fostering a Dynamic Global Economy* (Jackson Hole, Wyo.: Federal Reserve Bank of Kansas City, 2017), 115–28, https://ddorn.net/papers/Dorn-JacksonHole.pdf.

3. Wolfgang Dauth, Sebastian Findeisen, and Jens Suedekum, "Trade and Manufacturing Jobs in Germany," *American Economic Review* 107, no. 5 (2017): 337–42.

15985-06b_AutorCom&GD-3rdPgs.indd 474 5/12/22 12:06 PM

regions are not the same as those that were previously producing labor-intensive goods that directly competed with Chinese imports.

Hanson then turned to the example of Mexico, where there is evidence of moderate impacts of the China shock in regions that were specialized in labor-intensive manufacturing. He also noted that—along with Mexico—Hungary, Poland, Turkey, and Thailand were among the most exposed countries to the China trade shock given their pattern of comparative advantage. Hanson suggested that one key contributor in making the Mexican case different from other countries is that at the same time the China shock was wiping out labor-intensive manufacturing, the expansion of Mexican production chains in aerospace, automobiles, and medical devices took up the slack.[4] In Mexico, labor-intensive manufacturing and capital-intensive manufacturing are both located in the northern region of the country, while in the United States and Germany these sectors of manufacturing are located in different places. This lessened the aggregate impact in Mexico as the country's southern regions were not substantially affected and less regional reallocation was necessary. Hanson also noted that the implementation of NAFTA also occurred during this period and helped to boost Mexican manufacturing exports to the United States.

In regard to the Great Recession and the literature on the housing construction boom helping to mask the employment effects of the decline of manufacturing, Hanson agreed that housing trends across US regions played a role in their divergence.[5] Specifically, regions subject to the China shock largely avoided the subprime mortgage boom and instead saw much lower price appreciation in the early 2000s, causing the post-2006 collapse in construction to be much less significant for these places. Hanson also noted that many other elements could have made the Great Recession an important factor influencing the nature of adjustment, adding the example of the shutdown of credit as an important vehicle. He then raised the concept from Davis and Haltiwanger that depressed housing market values can also depress entrepreneurial capital for start-up firms, which could also be part of this story.[6]

4. Juan S. Blyde, Matias Busso, and Dario Romero Fonseca, "Labor Market Adjustment to Import Competition: Long-Run Evidence from Establishment Data," working paper 01100 (Washington: Inter-American Development Bank, 2020).

5. Kerwin Kofi Charles, Erik Hurst, and Matthew J. Notowidigdo, "The Masking of the Decline in Manufacturing Employment by the Housing Bubble," *Journal of Economic Perspectives* 30, no. 2 (2016): 179–200.

6. Steven Davis and John Haltiwanger, "Dynamism Diminished: The Role of Housing Markets and Credit Conditions," working paper 25466 (Cambridge, Mass.: National Bureau of Economic Research, 2019).

David Dorn then mentioned that several weeks following this event, the Institute for Fiscal Studies will release a chapter on trade that Dorn is coauthoring, which will include substantial discussion of the international differences between the impact of trade shocks on the United States and different European countries.[7] He also commented that a primary element is that the magnitude of exporting to China varies greatly across countries, while imports from China vary by a smaller degree. Germany and Switzerland are examples of the few Western countries that have a relatively even trade balance with China, while many others, like the United Kingdom, look exactly like the United States. An additional point he raised is that when looking at labor market impacts across countries in terms of manufacturing employment loss per unit of import competition shock, the United States does not look especially bad.[8] Some European countries, for example, Spain, have seen employment effects that are much more negative. He then highlighted several factors that explain some of this difference: labor market institutions and active labor market policies, initial industrial specialization, and exports to China. While some of these factors have traction, Dorn expressed his uncertainty that all factors contributing to these country divergences can be fully parsed out.

Finally, James Stock hypothesized that the authors could use their existing work to construct an energy transition shock, similar to the construction of the China shock, since there would be a sequence over time with regional specificity. In response, Hanson confirmed his interest in using earlier shocks to better understand the dimensions and elements of resilience, mentioning a recent grant proposal for a similar topic.

7. David Dorn and Peter Levell, "Trade and Inequality in Europe and the US," discussion paper 14914 (Bonn: Institute of Labor Economics, 2021), forthcoming in *IFS Deaton Review on Inequality*.

8. See figures 3 and 5 in Dorn and Levell, "Trade and Inequality."

15985-06b_AutorCom&GD-3rdPgs.indd 476 5/12/22 12:06 PM